# THE EFFECTS OF VISUALIZATIONS AND SPATIAL ABILITY ON LEARNING FROM STATIC MULTIMEDIA INSTRUCTIONS

A Dissertation
Presented to
The Academic Faculty

By

Keith A. Kline

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Psychology

Georgia Institute of Technology

August, 2012

**THE EFFECTS OF VISUALIZATIONS AND SPATIAL ABILITY ON LEARNING FROM STATIC MULTIMEDIA INSTRUCTIONS**


Approved by:

Dr. Richard Catrambone, Advisor
School of Psychology
*Georgia Institute of Technology*

Dr. Phillip Ackerman
School of Psychology
*Georgia Institute of Technology*

Dr. Frank Durso
School of Psychology
*Georgia Institute of Technology*

Dr. Wendy Rogers
School of Psychology
*Georgia Institute of Technology*

Dr. John Stasko
College of Computing
*Georgia Institute of Technology*



Date Approved: March 28, 2012

## ACKNOWLEDGEMENTS

I would like to thank my advisor, Richard Catrambone, and my fellow students in the Problem Solving and Educational Technology Lab for their advice and support. I have learned a great deal from each of them. I would also like to thank the members of my advisory committee for their guidance and careful review of my work. I also thank Dr. Jim Roberts who, although not on my committee, provided invaluable guidance for my statistical analysis.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... iii

LIST OF TABLES ..... vii

LIST OF FIGURES ..... ix

SUMMARY ..... xi

CHAPTER 1: INTRODUCTION ..... 1

1.1 Learning Assessment ..... 5
1.1.1 Structural and Dynamic Information ..... 5
1.1.2 Structural-Behavior-Function and Recall-Transfer are not Orthogonal. 7
1.2 Theories of Mental Representation ..... 8
1.2.1 Dual Coding Theory ..... 8
1.2.2 Cognitive Theory of Multimedia Learning ..... 10
1.2.3 Integrated Model of Text and Picture Comprehension ..... 11
1.3 Split Attention Impedes Learning ..... 12
1.4 Representational and Reasoning Strategies ..... 14
1.5 Supporting the Construction of Dynamic Mental Models ..... 17
1.5.1 Arrows Depict Movement ..... 20
1.5.2 Effects of arrows for low-spatial and low-VSWM learners ..... 20
1.5.3 Phase Diagrams with Arrows or Predictions ..... 22
1.6 Cognitive Abilities might Affect the Efficacy of Depictions ..... 23
1.6.1 Spatial ability ..... 24
1.6.2 Visuospatial Working Memory ..... 26
1.6.3 Similarities Between Spatial Ability and VSWM Capacity ..... 28
1.6.4 Discriminant Validity ..... 29

CHAPTER 2: EXPERIMENT 1 ........ 30

2.1 Method ........ 31
2.1.1 Participants ........ 31
2.1.2 Materials ........ 33
2.1.3 Design ........ 40
2.1.4 Procedure ........ 42
2.1.5 Scoring ........ 43
2.2 Results and Discussion ........ 44
2.2.1 Effects of Instructional Format, Topic, and Question Type ........ 45
2.2.2 Two-way Interactions ........ 49
2.2.3 Interaction among Topic, Format, and Question Type ........ 51
2.2.4 Individual Differences: Descriptive Statistics and Correlations ........ 53
2.2.5 Moderating Effect of Spatial Ability ........ 57
2.2.6 Moderating Effect of Visuospatial Working Memory ........ 61
2.2.7 Strategy use during Predictions ........ 62
2.3 Conclusion ........ 64

CHAPTER 3: EXPERIMENT 2 ........ 67

3.1 Methods ........ 68
3.1.1 Participants ........ 68
3.1.2 Materials, Procedure, and Design ........ 68
3.2 Results and Discussion ........ 71
3.2.1 Subjective Ratings of Attempts to Visualize Motion ........ 72
3.2.2 Effects of Instructional Format, Topic, and Question Type ........ 73
3.2.3 Interactions among Format, Topic, and Question Type ........ 76
3.2.4 Individual Differences: Descriptive Statistics and Correlations ........ 77
3.2.5 Individual Differences: Moderating Effects ........ 80
3.3 Conclusion ........ 83

CHAPTER 4: GENERAL DISCUSSION ........ 85

4.1 Acquisition of Behavioral and Structural Knowledge from Static Diagrams ........ 86

4.2 Environmental Support for Learning ...................................................... 87
4.3 Benefits of Making Predictions................................................................ 90
4.4 Active Learning from Optimal Instructions ............................................ 93

APPENDIX A: BACKGROUND QUESTIONNAIRE ............................................ 95

APPENDIX B: LESSON MATERIALS FOR THE PUMP TOPIC - SINGLE DIAGRAM CONDITION ............................................................................................ 96

APPENDIX C: LESSON MATERIALS FOR THE PUMP TOPIC – MULTIPHASE DIAGRAMS ........................................................................................................... 101

APPENDIX D: LEARNING ASSESSMENT QUESTIONS FOR THE PUMP TOPIC .................................................................................................................. 108

APPENDIX E: LESSON MATERIALS FOR THE PUMP TOPIC – MULTIPHASE DIAGRAMS WITH ARROWS.................................................................... 112

APPENDIX F: TEXT PASSAGE FOR THE CLOZE TEST ................................ 116

REFERENCES ......................................................................................................... 121

## LIST OF TABLES

Table 1. Instructional condition and topic assignments for participants 1-18 in Experiment 1. . 41

Table 2. Cronbach's alpha coefficients for the learning assessment questions in Experiment 1. 45

Table 3. Mean (SD) percent correct on the learning assessment questions from Experiment 1. 46

Table 4. Results for the post-hoc paired t-tests on Formats. 48

Table 5. Results for the post-hoc paired t-tests on the marginal means of Topic. 49

Table 6. Results of simple main effects analyses for the interaction between Topic and Question Type. 50

Table 7. Results of simple main effects analyses for the interaction between Topic and Format. 51

Table 8. Results for the paired t-tests on Instructional Formats within the Behavior questions on the Carburetor topic. 53

Table 9. Means and standard deviations for the measures of individual difference. Maximum possible scores are shown in parentheses. 53

Table 10. Split-half reliability estimates of individual differences tests in Experiment 1. The *r*-values represent correlations between the first and second half of each test. 54

Table 11. Correlations between standardized measures of individual differences in Experiment 1. 55

Table 12. Correlations between Composite individual difference measures. 55

Table 13. Descriptive statistics for pre-test performance in Experiment 1. 56

Table 14. Descriptive statistics for subjective reports of knowledge on the topics in Experiment 1. Responses ranged from 1-10. 56

Table 15. Correlations between pre-test scores and subjective knowledge ratings for the three topics in Experiment 1. 56

Table 16. Results for comparisons between hierarchical models including spatial ability and verbal ability. 58

Table 17. Results of the mediation analysis. Prediction performance mediated a significant amount of spatial ability's effect on learning. 60

Table 18. Results for comparisons between hierarchical models including visuospatial working memory (VSWM) and verbal ability. ........................................................................... 61

Table 19. Correlations between individual difference measures and percentage of propositional statements from participants' retrospective reports of strategy. ................................................. 63

Table 20. Instructional conditions for Experiment 2. ............................................................... 69

Table 21. Condition assignment and orders for the first 12 participants in Experiment 2. Multiphase, Simulation, and Arrows Formats were coded as 1, 2, and 3, respectively. ............. 70

Table 22. Internal consistency for the learning assessment in Experiment 2. ............................ 71

Table 23. Subjective ratings of extent to which participants attempted to visualize motion in diagrams. ................................................................................................................................ 72

Table 24. Mean (SD) percent correct on the learning assessment questions from Experiment 2. ............................................................................................................................................ 75

Table 25. Results for the post-hoc paired t-tests on the marginal means of Topic. ..................... 76

Table 26. Results for simple main effects analysis of the interaction between Topic and Question Type. ...................................................................................................................................... 76

Table 27. Means and standard deviations of the measures of individual differences in Experiment 2. ........................................................................................................................... 77

Table 28. Split-half reliability estimates of individual differences tests in Experiment 2. The r-values represent correlations between the first and second half of each test. .............................. 77

Table 29. Correlations between standardized measures of individual differences in Experiment 2. ............................................................................................................................................. 78

Table 30. Correlations between composite measures of individual differences in Experiment 2. ............................................................................................................................................ 79

Table 31. Descriptive statistics for pre-test performance in Experiment 2. ................................ 79

Table 32. Descriptive statistics for subjective reports of knowledge on the topics in Experiment 2. Responses ranged from 1-10. ........................................................................................... 79

Table 33. Correlations between objective and subjective measures of background knowledge on the three topics in Experiment 2. ......................................................................................... 80

Table 34. Results for comparisons between hierarchical models including spatial ability and verbal ability in Experiment 2 ................................................................................................. 81

Table 35. Results for comparisons between hierarchical models including VSWM and verbal ability in Experiment 2 ............................................................................................................ 83

## LIST OF FIGURES

Figure 1. Depiction of Paivio's dual-coding theory. Note that referential connections do not exist for some logogens and imagens; these are abstract words and nameless images.......................... 9

Figure 2. Mayer's (1997, 2001) cognitive theory of multimedia learning (CTML). Adapted with permission from Mayer (1997)...................................................................................... 11

Figure 3. The integrated model of text and picture comprehension. Adapted with permission from Schnotz and Bannert (2003)......................................................................................... 11

Figure 4. An example of test items from the Mental Rotations Test (Vandenberg & Kuse, 1978). ......................................................................................................................................... 37

Figure 5. A sample stimulus for the span task that is performed during the symmetry span test. This depicts the illumination of a single square that would be followed by other squares in the sequence.......................................................................................................................... 38

Figure 6. A sample stimulus for the symmetry judgment task that is performed during the symmetry span task......................................................................................................... 38

Figure 7. The set of arrows used in the rotation span test......................................................... 39

Figure 8. Performance on the learning assessment in Experiment 1 for Structure questions (upper panel) and Behavior questions (lower panel). .............................................................. 47

Figure 9. Interaction between Topic and Instructional Format.................................................. 51

Figure 10. Effects of spatial ability and instructional format on post-test performance. Scores from the learning assessment and spatial ability tests were standardized. .................................. 59

Figure 11. Model for spatial ability's effect on learning via mediation by performance on the prediction questions. B-weights and p-values refer to the results of three separate, pairwise regression analyses that were performed prior to the mediation analysis. .................................. 60

Figure 12. Non-significant (p = .10) interaction between VSWM and instructional format in Experiment 1.................................................................................................................. 62

Figure 13. Performance on the learning assessment in Experiment 2 for Structure questions (upper panel) and Behavior questions (lower panel). Error bars represent 95% confidence intervals......................................................................................................................... 74

Figure 14. Non-significant interaction (p = .10) between spatial ability and instructional conditions in Experiment two for post-test performance............................................................ 82

Figure 15. Single diagram lesson for the dual-action air pump, page 1. .................................... 96

Figure 16. Single diagram lesson for the dual-action air pump, page 2. .................................... 97

Figure 17. Single diagram lesson for the dual-action air pump, page 3. .................................... 98

Figure 18. Single diagram lesson for the dual-action air pump, page 4. .................................... 99

Figure 19. Single diagram lesson for the dual-action air pump, page 5. .................................. 100

Figure 20. Multiphase diagram lesson for the dual-action air pump, page 1............................ 101

Figure 21. Multiphase diagram lesson for the dual-action air pump, page 2............................ 102

Figure 22. Multiphase diagram lesson for the dual-action air pump, page 3............................ 103

Figure 23. Multiphase diagram lesson for the dual-action air pump, page 4............................ 104

Figure 24. Multiphase diagram lesson for the dual-action air pump, page 5............................ 105

Figure 25. Multiphase diagram lesson for the dual-action air pump, page 6............................ 106

Figure 26. Multiphase diagram lesson for the dual-action air pump, page 7............................ 107

Figure 27. Multiphase diagram lesson with arrows for the dual-action air pump, page 1......... 112

Figure 28. Multiphase diagram lesson with arrows for the dual-action air pump, page 2......... 113

Figure 29. Multiphase diagram lesson with arrows for the dual-action air pump, page 3......... 114

Figure 30. Multiphase diagram lesson with arrows for the dual-action air pump, page 4......... 115

Figure 31. Multiphase diagram lesson with arrows for the dual-action air pump, page 5......... 116

Figure 32. Multiphase diagram lesson with arrows for the dual-action air pump, page 6......... 117

Figure 33. Multiphase diagram lesson with arrows for the dual-action air pump, page 7......... 118

## SUMMARY

Successful learning about physical systems is thought to depend on the development of a mental representation of the system's dynamic behavior, which constitutes a mental model, rather than only its static structure (e.g., Schnotz, 2005). When learners must generate dynamic mental models from static diagrams, learning might be promoted by encouraging learners to visualize motion based on those diagrams. However, mental models represent dynamic spatial information that might be difficult to construct for learners with lower spatial ability; such learners might benefit from instructional designs that support spatial reasoning, such as phase diagrams and depictive arrows. In Experiment 1, participants learned about air pumps, carburetors, and toilet tanks from single phase diagrams, multiphase diagrams, or multiphase diagrams followed by a prediction activity in which they predicted system behavior in novel situations. This prediction activity was expected to implicitly prompt mental visualization of motion. Learning in the latter condition (i.e., with the prediction activity) was significantly better than learning in the single phase condition but not the multiphase condition without predictions. In the prediction condition, the enhancing effect of spatial ability on learning outcome was partially mediated by performance in the prediction activity. The mediation suggested that high spatial ability helped participants to accurately visualize the systems as they made predictions, which contributed to better performance on the learning assessment. Experiment 1 assessed visualizations during the prediction activity, whereas Experiment 2 assessed visualizations during the lessons. In two conditions in Experiment 2, participants were explicitly prompted to visualize motion in the system while viewing

the lessons. Because learners with lower spatial ability were expected to have difficulty visualizing motion, arrows depicting motion were added in one condition. A baseline condition excluded the arrows and the prompt to visualize motion. In all three conditions, participants viewed multiphase diagrams followed by the prediction activity. Learning outcomes among the three conditions did not differ significantly: Depictive arrows and prompts to visualize motion did not appear to improve learning. Also, spatial ability did not interact with instructional condition. However, both spatial ability and subjective ratings of attempts to visualize motion were predictive of learning outcome. Overall, results from the two experiments indicated that participants with higher spatial ability were better able than participants with lower spatial ability to generate dynamic mental models from static instructions, particularly when they were implicitly prompted to reason about the system as they made predictions. Learners with lower spatial ability might need other forms of assistance for mental model generation, such as animated instructions.

# CHAPTER 1: INTRODUCTION

Multimedia instructions present information in multiple formats, such as text and pictures. Prior research has demonstrated that learning is often improved by multimedia instructions in comparison to single-format instructions (e.g., Mayer, 2001). The effectiveness of multimedia instructions depends on several design features as well as individual differences in spatial ability and visuospatial working memory capacity (Hoffler & Leutner, 2010; Kline & Catrambone, 2011; Mayer, 2001; Meneghetti, Gyselinck, Pazzaglia, & De Beni, 2009).

Multimedia instructions typically include depictive spatial representations, such as pictures or animations, which directly convey the analog structure of information to be learned. Almost invariably, learners with higher spatial ability learn from multimedia as well as or better than their cohorts with lower spatial ability (e.g., Hegarty, Kriz, & Cate, 2003; Kline & Catrambone, 2009; Plass, Chun, Mayer, & Leutner, 2003). This has been interpreted as evidence that spatial ability contributes to the ability to comprehend instructions that convey spatial information (Gyselinck & Tardieu, 1999; Mayer, 2001; Mayer & Sims, 1994). The present studies specifically examined learning about physical systems. The goal was to assess several learning strategies and instructional designs that might assist learners with lower spatial ability, thereby reducing the learning gap caused by spatial ability. Two learning strategies were examined: performing mental simulations while viewing the lesson materials and generating predictions about system behavior immediately after viewing the lesson materials. Also, two instructional design factors were also examined: the number of system phases depicted in the lesson materials and the presence of arrows that depict movement in diagrams. These design factors were expected to facilitate mental simulations.

Mental simulations represent dynamic system behavior. Theories of multimedia learning (Mayer, 1997; Schnotz, 2005) have contended that the development of a mental model, which represents dynamic behavior, is an essential process for successful learning. A mental model is an analog representation that mirrors the structure and behavior of a system (Gentner & Stevens, 1983; Kieras & Bovair, 1984; Payne, 1991). It can be thought of as a mental image that can be manipulated and animated, enabling mental simulations and inferences about how the system functions. It is therefore a dynamic representation, rather than a static mental image.

Because mental models can be used to make inferences by simulation, they are useful for transferring knowledge to novel problems (Halasz, 1984; Kieras & Bovair, 1984; Staggers & Norcio, 1993; Wilson & Rutherford, 1989). For example, Kieras and Bovair (1984) asked participants to learn to operate a control panel, which consisted of a knob, dial, and switch that controlled a "phaser" gun on a fictitious starship. The group of participants who were given a diagram depicting the connections and power flow were better able to infer novel, efficient operating procedures than those who were not given the diagram. Knowledge of the diagram and its power flow were considered to constitute a mental model.

In a study with a physical system, Butcher (2006) asked students to learn about blood flow in the human circulatory system from either text alone or illustrated text. Those who viewed illustrations were better able to infer the functional purpose of system components (e.g., heart valves). Kieras and Bovair (1984), Butcher (2006), and others (e.g., Halasz, 1984) have found that learners who possess mental models often out-performed those who did not on measures of comprehension and knowledge transfer.

When learners develop dynamic mental models while learning, they might employ processes of spatial visualization – a component of spatial ability (Carroll, 1993). It has been assumed that spatial ability affects comprehension of multimedia instructions that include depictions. Some empirical results have supported this assumption. For example, Mayer and Sims (1994) found that the synchronization of animation and narration improved learning for those with higher spatial ability more than those with lower spatial ability. The synchronized condition was compared to asynchronous presentation. The authors inferred that spatial ability helped learners develop an integrated understanding of the narration and animation when they were presented simultaneously. With static materials, Kline and Catrambone (2009) found that performance on the Cube Comparison test, which is a measure of spatial ability, interacted with instructional design; those who performed well on the Cube Comparison test benefited more from the addition of illustrations to text.

Improved comprehension by high-spatial learners might be attributable to their ability to generate more accurate mental models. Some multimedia learning studies (e.g., Munzer, Seufert, & Brunken, 2009) have found that spatial ability affected knowledge of system processes and behavior (i.e., dynamic information that is represented in a mental model) but not system structure, implying that spatial ability affects mental models more than static mental images.

The present experiments evaluated several instructional design manipulations for their effectiveness in supporting mental model formation, particularly for learners with lower spatial ability or visuospatial working memory capacity (VSWM). Four instructional design factors were evaluated. Two of these factors can be classified as manipulations of external visualizations: (a) the number of system phases that were depicted, and (b) the inclusion of arrows depicting system behavior. The remaining two design factors can be classified as prompts that might have encouraged learners to develop a dynamic mental model while learning. Implicit prompts were given by asking participants to answer a few inference questions while viewing system diagrams (e.g., for an air pump, "What would happen if air leaked out through a hole in the upper passageway?). The questions demanded reasoning about motion in the system. Explicit prompts were given by instructing participants to attempt to visualize motion in the diagrams.

It was expected that multiphase diagrams and the prediction activity, when compared to the single phase diagrams, would enhance learning for participants with higher spatial ability and would provide less benefit for those with lower spatial ability. This pattern of enhancement would indicate that spatial ability conferred the necessary cognitive resources to benefit from the multiphase diagrams and the prediction activity. Specifically, spatial ability was expected to enable participants to better envision system transformation between multiphase diagrams. Similarly, spatial ability was expected to facilitate visualizations of movement while answering the prediction activities, which would lead to better mental model formation.

In contrast, it was expected that spatial ability would compensate for the lack of arrows depicting movement. This pattern of compensation would indicate that external cues (i.e., arrows) are needed by learners with lower spatial ability more than those with higher spatial ability.

It is important to note that *lower* and *higher* cognitive abilities are relative terms that denote relationships among cohorts in a given study. Furthermore, *lower* and *higher* are not intended to reflect a dichotomous construct. Cognitive abilities were treated as continuous variables in all analyses in the present experiments.

## 1.1 Learning Assessment

Oftentimes a goal of instructions on a physical systems is to enable the learners to gain a conceptual understanding of the system and transfer that understanding to novel situations (Betrancourt, Dillenbourg, & Clavien, 2008; Kim & Forbus, 1993; Mayer & Gallini, 1990). In contrast, verbatim memory and recall of non-conceptual information are less indicative of meaningful learning (Mayer & Gallini, 1990). Several studies have found that instructional manipulations are effective on measures of transfer but not recall (e.g., Catrambone, 1995; Kieras & Bovair, 1984; Mayer, 1981; Mayer & Gallini, 1990; Smith & Goodman, 1984). A possible reason for this dissociation might be that various instructional conditions can convey factual information equally well, but might differentially support the inferences that are needed for transfer.

In the present study, "learning" is used to refer to performance on a post-test. This measure is related to, but does not directly measure, an increase in knowledge from pre-test to post-test.

### 1.1.1 Structural and Dynamic Information

In addition to the distinction between knowledge recall and knowledge transfer, knowledge can be categorized according to structure-behavior-function theory (Bhatta &

Goel, 1997; Heiser & Tversky, 2006; Liu & Hmelo-Silver, 2009). Structural knowledge represents the names and spatial relations of system components. Behavioral knowledge represents causal relationships among components, including their movement. Functional knowledge represents the purpose of the behaviors. Thus, the categories can be viewed as a hierarchy, wherein structural knowledge is based on behavioral knowledge, which in turn is based on functional knowledge (Liu & Hmelo-Silver, 2009).

Structural information represents static relationships, whereas behavioral information represents dynamic movement. These are analogous to a mental image and a mental model, respectively. Mental images have been shown to share the analog properties of their real-world referents (Kosslyn, Ball, & Reiser, 1978; Pinker & Kosslyn, 1978). Similarly, mental models have been described as small scale analog representations of dynamic systems (Gentner & Stevens, 1983; Kieras & Bovair, 1984).

Although mental representations reflect external systems, they may not be isomorphic to those systems. Some evidence has demonstrated that animations of mental models are not isomorphic to the physical processes they represent. Physical systems often have many components that move simultaneously. Hegarty (1992) found that participants' eye movements and reaction times were suggestive of piecemeal animation of a pulley system. Rather than animating all components simultaneously, participants appeared to work sequentially through the causal chain.

In summary, structural knowledge is often acquired directly from instructions, and might be represented by a mental image that is isomorphic to the instructional image. In contrast, dynamic information must be inferred from static diagrams. Animations can convey dynamic information directly, but they are not temporally isomorphic to the mental models that they engender.

### 1.1.2 Structural-Behavior-Function and Recall-Transfer are not Orthogonal

Most prior research on instructional design has categorized questions on the recall-transfer dimension rather than the structure-behavior-function dimension. It is unlikely that structure questions require knowledge transfer, simply because structural information is often directly depicted or described, which precludes transfer questions about structural information. Thus, the recall-transfer dimension is not orthogonal to the structure-behavior-function dimension.

Questions that might implicitly evoke mental animations could be categorized as behavior questions. For example, the following two transfer questions used by Mayer and Gallini (1990) to assess learning about a car braking system could be categorized as behavior questions: “Suppose that you press on the brake pedal in your car, but the brakes don’t work. What could have gone wrong?” and “What happens when you pump the brakes?” For each question, most learners might envision the action and its consequences. Some researchers have argued that questions can be categorized *a priori* on the basis of whether they require static structural information about system components or dynamic information about movement (Heiser & Tversky, 2006; Liu & Hmelo-Silver, 2009). Munzer et al. (2009), who took this approach, found that performance on structure questions was not affected by spatial ability or instructional manipulations, whereas performance on behavior questions was. These results were congruent with the suggestion that structure questions typically require recall, which is often not affected by instructional manipulations, whereas behavior questions often require knowledge transfer, which is often affected by instructional manipulations.

## 1.2 Theories of Mental Representation

Both internal and external representations of information can take several forms. Certain information can be represented in either lexical or depictive formats. For example, the lexical representation, "the yellow house has six windows" can also be represented as an image. Dual-Coding theory (Paivio, 1986), related theories (Mayer, 2001; Schnotz, 2005), and the preceding example suggest that internal representations can be recoded into an alternate format. A reader might first create a lexical representation of the information about the yellow house and subsequently transform this into a mental image or vice versa. When learning about a physical system from text alone, many learners appeared to generate internal visualizations of the system (Gyselinck & Tardieu, 1999). Because the process generating visualizations is cognitively demanding and prone to error, the addition of illustrations to text sometimes improves comprehension. Learners can simply internalize pictorial representations instead of generating them (Gyselinck & Tardieu, 1999; Narayanan & Hegarty, 2002). Several compatible theories provide further explanation for the beneficial effect of illustrations.

### 1.2.1 Dual Coding Theory

Paivio (1986) developed dual coding theory, which posited that mental representations exist in either verbal or nonverbal format, corresponding to lexical and visuospatial representations, respectively (see Figure 1). Verbal information is directly encoded into a lexical representation system comprising "logogens," whereas pictorial and visuospatial information is directly encoded into a nonverbal representation system comprising "imagens." Associative networks exist within each of these systems (e.g., a semantic network exists within the verbal channel). A key claim of dual-coding theory lies in the referential connections between the two systems, whereby information can be recoded and retrieved bidirectionally.

Dual coding theory has inspired at least two models of the cognitive processes that occur during learning from multimedia materials. These are Mayer's (2001) cognitive theory of multimedia learning and Schnotz's (2005) integrated model of text and picture comprehension.

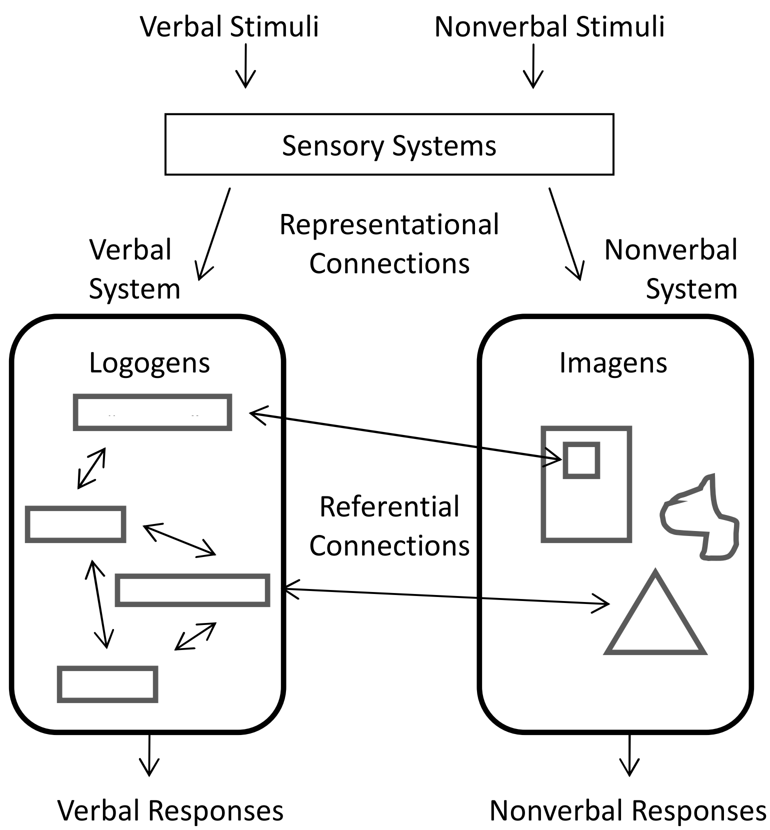


*Figure 1.* Depiction of Paivio's dual-coding theory. Note that referential connections do not exist for some logogens and imagens; these are abstract words and nameless images.

In both models, learning about physical systems from multimedia is construed as a process of mental model construction during which visual and verbal representations are integrated (see also Narayanan & Hegarty, 2002).

Both models share the basic two-channel structure proposed by dual coding theory. The models are very similar, but differ in emphasis and several details (Dutke & Rinck, 2006; Gyselinck & Tardieu, 1999). Mayer (2001) emphasized the role of sensory modality, and maintains separation between the two processing channels, whereas Schnotz (2005) disregarded modality and emphasized the processes by which information from the two channels is integrated.

### 1.2.2 Cognitive Theory of Multimedia Learning

Mayer's cognitive theory of multimedia learning (CTML; Mayer, 2001; Mayer & Moreno, 2003) builds upon dual coding theory, placing emphasis on the beneficial effects of the active integration of verbal and pictorial information. The model (see Figure 2) implies that mental models are constructed separately in the two processing channels, and integration occurs only at the final stage (Dutke & Rinck, 2006). The integration of verbal and visuospatial representations is thought to be an active, effortful process that improves comprehension and retention. Notably, the model does not culminate in a single mental model abstracted from the two channels.

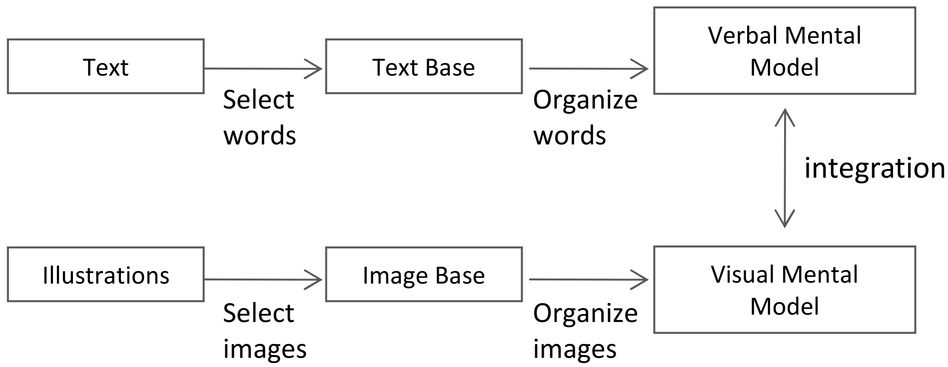


*Figure 2.* Mayer's (1997, 2001) cognitive theory of multimedia learning (CTML). Adapted with permission from Mayer (1997).

### 1.2.3 Integrated Model of Text and Picture Comprehension

In contrast to the CTML, the integrated model of text and picture comprehension (Schnotz, 2005), proposes that a single, unified mental model arises from intermediate verbal and visuospatial representations (see Figure 3). Moreover, communication between the two channels occurs at all stages of the learning process—denoted by the diagonal and vertical arrows in Figure 3.

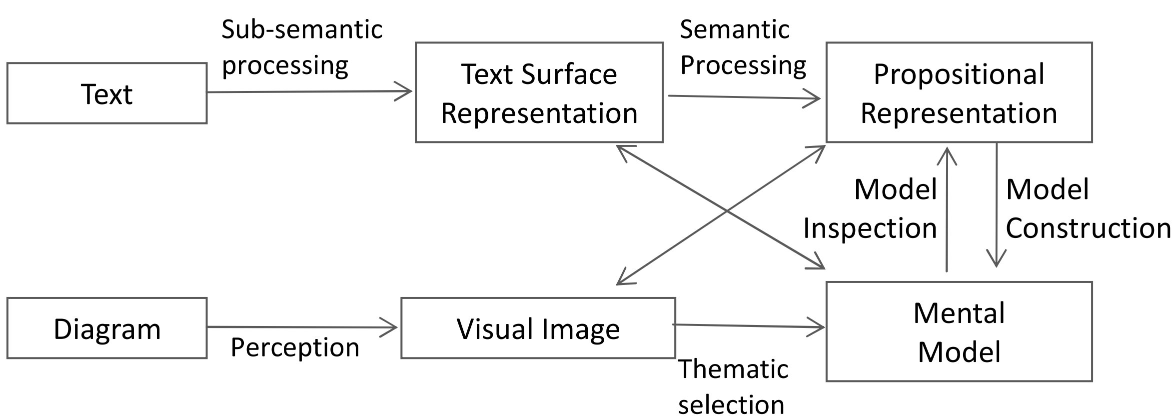


*Figure 3.* The integrated model of text and picture comprehension. Adapted with permission from Schnotz and Bannert (2003).

The CTML and the integrated model of text and picture comprehension are quite similar, as can be seen from a comparison of Figure 2 and Figure 3. An important point of divergence is that the two channels communicate with each other throughout the comprehension process in the integrated model of text and picture comprehension, whereas the two channels communicate only at the final stage in the CTML (Dutke & Rinck, 2006). Although Mayer (1997, 2001) did not explicitly state that this separation was intended, it seems to be implied by the graphical depiction and verbal descriptions of the model. That is, the graphical depiction of the model (see Figure 2) contains no connections between verbal and spatial channels until the final integration stage. Mayer's description of the model corroborates the depicted separation of channels. However, Mayer's (Mayer & Sims, 1994) earlier findings suggested that integration between the channels occurs before the learning session is completed. Specifically, simultaneous presentation of verbal and pictorial information improved learning when compared to successive presentations, which suggested that learners were integrating the simultaneously presented verbal and pictorial information throughout the learning session. It is possible that this integration occurs iteratively throughout the learning process, rather than only once at the end of the process.

In summary, the integrated model of text and picture comprehension (Schnotz & Bannert, 2003) and empirical results (Mayer & Sims, 1994) suggest that descriptive and depictive representations should complement each other when they are presented simultaneously. However, the presence of multiple sources of information introduces the problem of appropriately switching attention between the sources.

## 1.3 Split Attention Impedes Learning

All instructions in the present experiments were designed to limit split attention between text and diagrams. The presentation of information in two complementary formats (e.g., text and diagrams) creates the potential problem that attention might be split between the two sources. Oftentimes a learner must inspect a diagram to understand a text passage and vice versa. This requires many shifts of attention between the two sources of information (Ayres & Paas, 2007; Betrancourt, 2005; Hegarty & Just, 1993). Moreover, visual search is required to locate corresponding diagram components and text segments (Ginns, 2006). Attention shifts and visual searches are not thought to contribute to schema acquisition (Chandler & Sweller, 1991). Rather, these activities deplete cognitive resources that would otherwise be available for learning.

It is widely accepted that working memory is a limited resource (e.g., Baddeley, 2002). It is also believed that working memory is used heavily during the learning process to construct new knowledge and to integrate new information with existing knowledge (Gyselinck, Ehrlich, Cornoldi, de Beni, & Dubois, 2000; Mayer & Moreno, 2003). Therefore, instructional designs that place extraneous load on working memory should impede learning.

Cognitive load theory posits that working memory resources are distributed among three sources of load, including intrinsic, extraneous, and germane load (Paas, Renkl, & Sweller, 2003; Sweller, 2010). Intrinsic cognitive load arises from the complexity of the content; complexity is defined as the number of interacting elements in the system that must be considered simultaneously to understand the action of any given element. Extraneous cognitive load arises from inefficient instructional designs that require learners to exert effort that is not directly associated with schema acquisition (e.g., visual search). Germane cognitive load arises from cognitive processes that are beneficial for learning, such as making inferences. Instructional designers should seek to minimize extraneous cognitive load so that resources are available for germane processes.

Extraneous cognitive load can be reduced by the use of color codes or numerical labels to match portions of text with corresponding picture elements (Florax & Ploetzner, 2010; Kalyuga, Chandler, & Sweller, 1999). The spatial integration of text with pictures has also been found to reduce extraneous cognitive load (Cierniak, Scheiter, & Gerjets, 2009). Cierniak et al. asked participants to learn about and the anatomy and biochemistry of the kidney. Expository text and names of anatomical parts were presented either separately from the picture (in a paragraph) or integrated with the picture, such that corresponding descriptions and names were immediately adjacent and connected by lines. Learning outcomes improved and subjective ratings of cognitive load decreased in the integrated format. The present experiments employed the spatial contiguity principle in text placement, and additionally showed lines that connected keywords in the text to their corresponding diagrammatic elements. Thus, the instructions were designed to enable learners to develop a mental model that included both verbal and visuospatial information. From these verbal and visual representations, learners should be able to reason about how the system would operate in novel situations (i.e., knowledge transfer).

## 1.4 Representational and Reasoning Strategies

It is possible to reason about a system by using propositional representations rather than visuospatial mental models. Participants with high spatial ability might select a visuospatial strategy more often than participants with low spatial ability who might select a verbal strategy more often. This would occur if learners tend to select a strategy that matches their abilities.

Researchers have identified reasoning and representational strategies in a variety of tasks with several metrics. In syllogistic reasoning tasks, participants are given two premises and are asked to generate a conclusion. An example of a syllogism follows:

Premise 1: All bankers are fastidious

Premise 2: Some females are bankers

Conclusion: Some females are fastidious

Syllogisms can be solved with a verbal or visual strategy. The verbal strategy maintains the propositional representations of the premises, whereas the visual strategy converts the premises into Euler circles (i.e., [non]intersecting circles). When asked to describe how they solve syllogisms, participants demonstrated consistent strategy selection (Bacon, Handley, & Newstead, 2003; Ford, 1995). The verbal-strategy participants drew circles around words in the premises and connected the words with lines and notes to represent relationships. The spatial-strategy users drew Euler circles to represent the relationships.

Transitive inference tasks, also called linear syllogisms, require relational reasoning about three items. The relations among the three items can be described on a single dimension (e.g., size); for example, A is larger than B, and B is larger C. Participants would be asked about the relation between A and C. Egan and Grimes-Farrow (1982) identified two strategies that were used by participants. Spatial-strategy users reported ordering items on an imagined axis (e.g., placing lighter and heavier items on the left and right ends of a horizontal axis, respectively). Visual-strategy reported generating mental images of the items.

Representational strategy selection has also been studied in the sentence-picture verification task (Macleod, Hunt, & Mathews, 1978). In this task, participants determine as quickly as possible whether a picture matches a previously displayed sentence. Each trial begins with a simple sentence, such as "The star is above the plus." When ready, the participant presses a key and a picture is displayed that either matches the sentence or does not match (e.g., a picture of a star above a plus sign). The participant reports as quickly as possible whether the picture matches the sentence.

To make the comparison, the sentence and picture must be converted into a common format. On any given trial, a participant can adopt one of two strategies. He or she can convert the sentence to a pictorial representation before pressing the space bar to view the picture, or instead might maintain a verbal representation of the sentence until the second picture is displayed. The strategy most commonly adopted by a participant can be deduced by examining reaction times after the picture is displayed.

If a verbal strategy is chosen, then certain linguistic properties of the sentence should affect reaction time. Specifically, sentences that include negation, such as "the star is not above the plus" should require additional cognitive steps when using the verbal strategy (Carpenter & Just, 1975). If a spatial strategy is adopted, this cognitive step is accomplished during encoding of the sentence, which means that reaction time to the picture should be unaffected by negation. Based on this rationale, Macleod et al. (1978) were able to identify two groups of participants who used either a verbal or visual encoding strategy. Macleod et al. also found evidence that cognitive abilities affected reaction time. For those who used a verbal strategy, verbal ability was highly correlated with reaction time. For those who used a spatial strategy, spatial ability was highly correlated with reaction time. Furthermore, spatial ability was significantly higher for the group that exhibited the spatial strategy.

The three paradigms discussed above are qualitatively different than reasoning about mechanical systems. Nonetheless, they demonstrate how different participants might select different strategies in various reasoning tasks. There is very little literature that addresses how experimenters can identify participants' selection of strategies in the context of reasoning about physical systems. Williams, Hollan, and Stevens (1983) recorded verbal protocols of participants while they answered questions about a simple heat exchanger after learning from diagrams and text,. Participants used many phrases that referred to spatial relations of system components (e.g., "passes through," and "comes in"), yet such phrases do not distinguish between the use of propositional or visuospatial representations. Schwartz and Black (1996) reported that participants often used hand motions when they reasoned about a gear system that had been verbally described; the hand motions were taken as evidence that participants were animating a mental image. In the present experiments, an attempt was made to identify strategy selection by collecting written protocols after participants answer transfer questions. Correlations between strategies and cognitive abilities were computed, and the effect of strategy on performance was assessed.

### 1.5 Supporting the Construction of Dynamic Mental Models

Perhaps the most direct way of facilitating the construction of a dynamic mental model would be to present dynamic animations. One benefit of animations is that transitions between system phases do not need to be inferred by the learner, because transitions are overtly depicted (Hoffler & Leutner, 2010). Hoffler & Leutner (2006,

2010) found that animations were particularly beneficial for learners with lower spatial ability. Spatial ability was significantly more predictive of learning outcome from static phase diagrams than from animation. The authors inferred that high spatial ability helped some learners to visualize transformations between the phase diagrams. Learners with lower spatial ability were presumably less able to do this, and therefore benefited more from animations.

Despite the potential benefits of animations for conveying dynamic information about physical systems, they carry a number of practical and theoretical limitations. Practically, they are often more costly and time consuming to produce than static illustrations, and they require more technical expertise from the instructional designer. For learners, animations might be difficult to use as a reference after initial learning. Rather than simply finding a desired page in a textbook or electronic document, a learner would have to fast-forward/rewind to the desired place in the animation (although chapter marks might help to some extent). Moreover, research has demonstrated several drawbacks of animations that arise from the cognitive demands of viewing and comprehending animations (for a review see Tversky, Morrison, & Betrancourt, 2002). First, learners must infer referential connections between the narration and animation (Mayer & Moreno, 2003). For example, when one hears “the inlet valve opens,” one must identify the inlet valve in the animation. This imposes extraneous cognitive load. The problem can be remedied by visual cues that direct attention to items as they are described in the narration (de Koning, Tabbers, Rikers, & Paas, 2007; Faraday & Sutcliffe, 1997; Jamet, Gavota, & Quaireau, 2008).

A second, related drawback of animation is that multiple changes often occur simultaneously in the display, and learners might have difficulty perceiving and comprehending them. For example, an air pump's inlet and outlet valves open simultaneously. Although this would be an accurate depiction of how the system works, it would not coincide with the piecemeal mental animation of physical systems exhibited by participants (Hegarty, 1992; Hegarty & Sims, 1994; Just & Carpenter, 1985). This might explain why Hegarty et al. (2003) observed better learning in a condition that used three phase diagrams than in a condition that used animation, and why animations have failed to improve learning in many studies (Tversky, et al., 2002).

Finally, animations might reduce cognitive processes that are germane to schema acquisition (Betrancourt, 2005; Mayer, Hegarty, Mayer, & Campbell, 2005; Schnotz & Lowe, 2003). When an image is animated on-screen, the learner simply has to encode the animation rather than generate his or her own mental animation. To the extent of that successful learning is an active, generative process (Chi, 2009; Mayer, et al., 2005), the construction of mental animations should enhance learning. The act of generating a mental animation while learning might facilitate subsequent mental animations that are demanded by transfer tests.

In summary, animated multimedia might be suboptimal because learners are (a) required to correctly identify visual elements that match the narration, (b) required to attend to multiple moving parts simultaneously, and (c) relieved of generating germane mental animations. These factors might explain the large number of studies that have failed to show benefit of animation over static diagrams (Ayres & Paas, 2007; Tversky, et al., 2002).

### 1.5.1 Arrows Depict Movement

A static diagram without motion cues can require learners to infer the direction of motion of system components (Betrancourt, 2005). Heiser and Tversky (2006) found evidence that arrows in diagrams can be an effective means of depicting motion. Participants were given a diagram that depicted a car brake, an air pump, or a pulley system. For half of the participants, the diagrams included arrows that depicted movement of components, fluid, and/or air. Below the diagrams, participants were asked to write a description of the system. Their descriptions were coded into a set of propositions, each of which was classified as conveying structural or behavioral (i.e., movement-related) information. Participants who saw arrows produced significantly more movement-related propositions and significantly less structural propositions than those who did not see arrows.

In a second experiment, Heiser and Tversky (2006) asked participants to produce diagrams after reading text passages that described the system and structural or functional terms. The latter group produced a significantly higher number of the arrows in their diagrams.

### 1.5.2 Effects of arrows for low-spatial and low-VSWM learners

When asked to verify whether a sentence and diagram of a pulley system were congruent, participants with lower spatial ability made more errors than those with higher spatial ability (Hegarty & Sims, 1994). The difference between low and high ability participants was attenuated when they were allowed to draw on the diagrams (Hegarty & Steinhoff, 1997). Those with low spatial ability appeared to rely heavily on drawing arrows on the pulley diagram, whereas those with high spatial ability drew on the diagrams significantly less. These results are consistent with the notion that external diagrams can act as a memory aid, which would be especially useful for learners with lower visuospatial working memory capacity.

Learners may need a memory aid if they work through a causal chain in a piecemeal fashion. Eye movement data have shown that participants worked through a causal system (a pulley system) in a piecemeal fashion, working progressively through the causal chain (Hegarty & Just, 1993). If participants are allowed to draw on diagrams, they need not hold information about a previous component in working memory while mentally animating a subsequent component.

The pulley and gear problems used by Hegarty and colleagues (Hegarty & Just, 1993; Hegarty & Sims, 1994; Hegarty & Steinhoff, 1997) are similar in some ways to the problems that constitute tests of spatial visualization ability. For example, in the surface development test (Ekstrom, French, Harman, & Dermen, 1976) participants see a depiction of a two dimensional object, and must perform successive mental folding operations to determine how the object would appear when folded into a three-dimensional object. Thus, it is not surprising that spatial ability would affect performance in the pulley and gear problems. It is notable, however, that participants with lower spatial ability performed well when they were allowed to draw arrows on the diagrams. This suggests that they were sufficiently able to comprehend and reason about the diagrams when arrows were present; their performance deficit might have been due to an inability to accurately animate a mental image without external movement cues.

### 1.5.3 Phase Diagrams with Arrows or Predictions

Differences between successive phase diagrams are implicit indications of movement. By comparing phase diagrams of a system in different states, learners might be able to infer how a system moves between those states. Unfortunately, experimental confounds were present in three prior studies that compared single phase diagrams to multiphase diagrams. Both Munzer et al. (2009) and Mayer and Gallini (1990) added arrows to multiphase depictions, and compared them to single phase depictions without arrows. Improved learning in the former condition might have been attributable to either the presence of arrows or the multiphase depictions.

Hegarty et al. (2003) also found that multiphase diagrams might improve learning, yet again there was an experimental confound. Participants who learned from instructions depicting three phases of a toilet tank performed better on transfer questions than participants who learned from a single-phase diagram. However, the multiphase condition was coupled with prediction questions during the learning session, thereby confounding the effects of viewing multiphase depictions and making predictions. While viewing the diagrams, learners were asked to predict how the system would react to certain conditions (e.g., “When the handle is pushed down, what happens to the upper and lower disks?” p. 334). The authors suggested that such prediction questions would prompt mental animations and improve mental model formation.

It is possible that the act of generating predictions during the learning phase could produce similar knowledge as that derived from viewing animations, provided that participants make correct predictions. Byrne, Stasko, and Catrambone (1999) found that learners who were asked to predict the behavior of a search algorithm from static diagrams learned as well as those who viewed animations of algorithm behavior without making predictions. Participants were given immediate feedback about the correctness of their predictions, which means that participants in the prediction condition might have received more information during the learning session. Nonetheless, it is possible that there was a beneficial effect of making predictions, regardless of feedback, while viewing diagrams.

The accuracy of learners' predictions about how a system works might depend on their cognitive ability. Some learners might be better able to make predictions than others, and they might therefore benefit more from the prediction activity. Thus, cognitive ability might moderate the effectiveness of a prediction activity.

## 1.6 Cognitive Abilities might Affect the Efficacy of Depictions

Prior research has shown that individual differences in spatial ability and visuospatial short-term memory moderate the effectiveness of certain multimedia design manipulations (Gyselinck, Cornoldi, Dubois, De Beni, & Ehrlich, 2002; Gyselinck, et al., 2000; Gyselinck & Tardieu, 1999; Hoffler & Leutner, 2010; Kline & Catrambone, 2009; Lee, 2007; Mayer & Sims, 1994; Meneghetti, et al., 2009; Plass, et al., 2003; Yang, Andre, & Greenbowe, 2003). In these studies, each group of researchers has addressed either visuospatial short-term memory or spatial ability, but not both. (Note that although Gyselinck and colleagues discussed visuospatial *working* memory, they actually measured visuospatial *short-term* memory). Gyselinck and colleagues have examined only visuospatial short-term memory, whereas all other researchers have examined only spatial ability. In addition to spatial ability, the present experiments assessed visuospatial working memory instead of short-term memory, because the executive control processes that underlie working memory – but not short-term memory – were expected to affect learning.

### 1.6.1 Spatial ability

Spatial ability is broadly defined as the ability to perceive, generate, briefly retain, and mentally manipulate spatial information (Carroll, 1993; Lohman, 1996). Spatial ability consists of multiple factors that can be differentiated by factor analysis of psychometric test performance (Carroll, 1993; see also Hegarty & Waller, 2005).

The information processing approach to psychometric test performance considers the various perceptual and cognitive processes that are demanded by each test (Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Hegarty & Waller, 2005; Just & Carpenter, 1985). This approach has revealed that individual differences in performance are the result of "differences in speed of encoding and transforming spatial information, spatial working memory capacity, and strategies" (Hegarty et al., 2006, p. 154). Different tests of spatial ability demand different processes; this explains why factor analytic studies (e.g., Carroll, 1993) reveal dissociable sub-factors.

On logical rather than empirical grounds, authors (Eliot & Smith, 1983; Stumpf & Eliot, 1999) identified two broad categories of spatial abilities: a recognition category and a manipulation category. The cognitive processes associated with the recognition category include stimulus perception, retention, and sometimes a manipulation on a single plane (i.e. two-dimensional rotation). The cognitive processes associated with the manipulation category include those associated with the recognition category in addition to the manipulation of objects in three dimensions.

Psychometric tests in the manipulation category can be considered more complex than those in the recognition category, because the former requires not only stimulus perception and retention, but also manipulation in three dimensions (Eliot & Smith, 1983). Tests in the recognition category typically consist of simple problems administered under high time pressure. Therefore, recognition and manipulation tests are considered to be measures of speed and power, respectively (Pellegrino, Alderton, & Shute, 1984).

Carroll's (1993) factor analysis of spatial ability confirmed the logical distinctions described by Eliot and Smith (Eliot & Smith, 1983) and Pellegrino et al. (1984). Tests requiring manipulations under low time pressure were found to load onto a single factor, which Carroll called visualization (Vz). Other tests administered under more speeded conditions loaded onto four different factors, including spatial relations, closure flexibility, closure speed, and perceptual speed.

Tests of spatial relations (SR) require two-dimensional translation and/or rotation. In the SR tests, entire images are mentally rotated as a whole. For example, a participant might be asked whether a mirrored and rotated letter *R* could be rotated to match a non-mirrored upright *R*. When the *R* is mentally rotated, the relations among the vertices and lines do not change.

The term *manipulation*, when used in reference to Vz tests, refers to much more than the simple two-dimensional translation or rotation that is needed for SR tests. Most Vz test items require manipulations that change relations *within* a mental image. For example, in the Surface Development Test (Ekstrom, et al., 1976), participants must visualize the folding of a flattened cardboard cutout to determine how it would appear when folded into a three dimensional object. With each mental fold, relations among vertices and edges change. Similarly, when participants visualize the movement of components in a physical system (e.g., a carburetor), the relations among system components change. Thus, some of the cognitive operations needed for learning from illustrations are very similar to those needed for Vz test performance, whereas those needed for SR test performance are less similar. Moreover, Vz tests and the learning conditions of the present study enforced little time pressure, whereas SR tests enforce high time pressure. For these reasons Vz but not SR was assessed in the present study.

### 1.6.2 Visuospatial Working Memory

Working memory is a multi-component system that is responsible for actively maintaining information while engaging in other related or unrelated tasks (Conway, et al., 2005). The system consists of a central executive, a phonological loop, and a visuospatial sketchpad (Baddeley, 2002). The latter two subsystems are dedicated to domain-specific storage (i.e., verbal and visuospatial information, respectively), whereas the central executive is dedicated to domain-general processing, such as attention allocation.

Working memory tasks are distinct from short-term memory tasks (Conway, et al., 2005; Kane, et al., 2004). Working memory tasks place high demand on the central executive, whereas short-term memory tasks simply require domain-specific storage and rehearsal. Typical working memory tasks present a series of items to be remembered; the items are interleaved with a secondary task such as arithmetic. These tasks are referred to as storage and processing components, respectively. The central executive is responsible for the processing task and alternating between the two tasks; the two slave systems are responsible for the storage task.

**Is working memory domain-general or domain-specific?** Although there is clear evidence for the dissociation between verbal and visuospatial storage (for a review see Baddeley, 2002), there has been debate about whether working memory should be decomposed into verbal and visuospatial components. In support of the dissociation, Shah and Miyake (1996) reported that performance on a visuospatial working memory task was correlated with visualization ability but not verbal SAT scores, whereas performance on a verbal working memory task was correlated with verbal SAT scores but not visualization ability.

However, results from another study (Kane, et al., 2004) suggested that Shah and Miyake's (1996) findings might be explained by domain-specific storage requirements. That is, the tasks that correlate with each other might draw on the same domain-specific storage mechanisms. Kane et al. used factor analyses and structural equation models to derive four latent variables, representing verbal working memory, verbal short-term memory, visuospatial working memory, and visuospatial short-term memory. They concluded that performance on working memory tests reflected domain-general working memory capacity and domain-specific storage capacity.

The present studies focused on comprehension of various types of graphics rather than various texts—the texts varied little among conditions. Therefore, visuospatial working memory was measured, but verbal working memory was not. When attempting to comprehend multiple, related phase diagrams, learners might temporarily store representations of one phase diagram as they visually inspect another so that comparisons can be made.  Thus, VSWM capacity was expected to affect learning in conditions containing multiphase diagrams.  Moreover, domain-general executive attention might be used to shift attention between diagrams and text.

Sweller and Chandler (1994) argued that VSWM is crucial for learning about physical systems, particularly when the systems contain multiple interacting elements (Sweller & Chandler, 1994).  VSWM capacity might be particularly important when learning from illustrated texts.  Executive control of attention, which underlies working memory, might be critical for switching attention between text and diagrams.  It might also be critical for comparing successive diagrams.  In the present experiments, it was expected that VSWM would moderate learning in all instructional conditions; all conditions included at least one diagram accompanied by text, and it was assumed that participants would shift attention between the text and diagrams.  In those conditions that contained multiple diagrams that participants might compare, VSWM was expected to be even more predictive of learning.

### 1.6.3  Similarities Between Spatial Ability and VSWM Capacity

Empirical evidence has shown considerable overlap between the constructs of VSWM and spatial ability (particularly Vz).  Miyake, Friedman, Rettinger, Shah, and Hegarty (2001) assessed the relationship among domain-specific storage capacities, executive function, and three subfactors of spatial ability, including visualization (Vz), spatial relations (SR), and perceptual speed (PS).  They concluded from structural equation modeling that performances on tests of the three spatial ability subfactors were significantly related to executive functioning (standardized factor loadings for Vz, SR, and PS onto executive functioning were .91, .83, and .43, respectively).

Similar evidence came from a meta-analysis of the relationships among intelligence, working memory, and cognitive abilities by Ackerman, Beier, and Boyle (2005). They estimated that the true score correlation between spatial ability and visuospatial working memory was .49.

Finally, a rational assessment of VSWM and Vz tests suggests that they demand some similar cognitive operations. Tests of VSWM require participants to store in working memory a series of visual items while performing another cognitive task, such as mental rotation or symmetry judgments. Likewise, many tests of Vz require participants to store in working memory a series of transitional states while generating new transitional states. For example, in the Surface Development test (Ekstrom et al., 1976), participants must successively fold a mental image; as each fold is performed, the result of previous folds must be maintained.

### 1.6.4 Discriminant Validity

A possible explanation for the effect of spatial ability and VSWM on multimedia learning is simply the fact that they are correlated with general intelligence, or "g" (Ackerman, et al., 2005; Kane, et al., 2004), which of course would affect learning. Verbal ability, which is also correlated with *g*, was also measured in the present experiments. Through hierarchical regression, variance associated with verbal ability was partialled out prior to assessing the effects of spatial ability and VSWM. Thus, variance common to *g*, verbal ability, and spatial ability was partialled out prior to examining the effect of spatial ability alone. A similar analysis was performed for VSWM.

## CHAPTER 2: EXPERIMENT 1

Experiment 1 assessed how learning about physical systems from multimedia instructions was affected by the following factors: the number of phase diagrams depicted in instructions, learners' attempts to answer prediction questions immediately after the learning session, individual differences in spatial ability, and individual differences in visuospatial working memory capacity (VSWM). Whereas previous studies (Hegarty, et al., 2003; Mayer & Gallini, 1990) have confounded the effects of viewing multiphase diagrams and answering prediction questions, these variables were manipulated separately here.

The multiphase diagrams depicted the systems in all of their major states (e.g., handle up, handle down for a pump, valves open, valves closed); these depictions were expected to reduce the need for participants to visualize non-depicted states. However, participants might still have benefited from visualizing transitions between the states to generate a complete mental model. Because spatial ability enables such visualizations, it was expected to be predictive of learning outcome in this condition. Participants with high spatial ability were expected to generate accurate mental models by visualizing transitions between depicted system states. Participants with lower spatial ability were expected to be less able to accurately visualize the transitions.

Several prior studies have found similar patterns of results—namely, that spatial ability enhances learning from improved instructional designs in comparison to less-than-optimal designs. Mayer's (2001) "individual differences principle" states that "design effects are stronger for high-spatial learners rather than for low-spatial learners" (p. 161, also see Huk, 2006; Lee, 2007). This pattern of enhancement by spatial ability has been found for a variety of instructional design manipulations, including the addition of illustrations to text-only instructions (Kline & Catrambone, 2009), the synchronization of narration and animation (Mayer & Sims, 1994), the use of three-dimensional rather than two-dimensional illustrations (Huk, 2006) and other manipulations (Plass, et al., 2003; Yang, et al., 2003). In light of these studies, it was expected that learners with higher spatial ability would be better able than those with lower spatial ability to benefit from the multiphase depictions.

Although the presentation of multiphase diagrams might encourage learners to generate a dynamic mental model, it is by no means guaranteed that they would do so. To increase the chance of accurate mental model formation, participants in one condition were asked to predict how the system would behave in various situations. It was expected that this activity would implicitly prompt mental animations (or "simulations"), which would result in better mental model formation and better performance on subsequent transfer questions. To control for the amount of information that participants received during the learning session, they were not given feedback about the correctness of their answers to these prediction questions.

Implicit prompting is not, of course, guaranteed to elicit mental animation. To determine whether the prompting was effective in this regard, participants were asked to retrospectively report how they answered the prediction questions (see below).

## 2.1 Method

### 2.1.1 Participants

Seventy-two undergraduate students (Male = 40, Female = 30, Unreported = 2) from the Georgia Institute of Technology participated in the experiment. They were recruited from the psychology student subject pool on a volunteer basis. Participants were compensated with credits, which course instructors typically applied as extra credit.

A power analysis was used to determine the sample size for the experiment. An estimated effect size of $f^2 = .057$ was used, based on the results of a previous experiment with a similar design (Kline & Catrambone, 2011). The power analysis showed that 75 participants would be needed for a 0.8 chance of detecting a moderating effect of spatial ability on the instructional manipulation. To achieve a counterbalanced design (see below), data from 72 rather than 75 participants was used. Seven participants did not return for the second experimental session, and their data was replaced by newly recruited participants. The participants chose not to return, but did not provide an explanation for withdrawing. Participants who did not return for the second session performed significantly worse on a composite measure of standardized spatial ability and verbal ability than participants who did return, $F(1,77) = 4.77$, $p = .032$, $\eta^2 = .06$. VSWM was not examined, because the data files were overwritten with replacement data. The significant difference in scores between returning and non-returning participants might have been due to low motivation: The seven participants might have been unmotivated to return and unmotivated to achieve high performance on the verbal and spatial ability tests. Also, their decision to withdraw might have been caused by frustration experienced during the tests.

During the piloting of a recent study (Kline & Catrambone, 2011), it was discovered that many participants from this subject pool exert very little effort on cognitive abilities tests and tests of knowledge retention. Many participants did not attempt to answer more than half of the learning assessment questions, and often skipped portions of the abilities tests. To address this problem in the present experiment, participants were offered a performance-based financial reward ($10 to the top five performers).

### 2.1.2 Materials

Each participant learned about three topics: carburetors, dual-action pumps (henceforth "pump"), and toilet tanks. For each participant, each topic was presented in a different instructional format (one of the three formats described below). Three tests of spatial ability, three tests of verbal ability, and two tests of VSWM were administered.

***Selection of topics.*** Multimedia instructional designs have been studied on a wide variety of learning topics. Most frequently, participants learned about physical processes, such as air pump action (Kline & Catrambone, 2009; Mayer & Sims, 1994), automobile brakes (Mayer & Gallini, 1990), gas laws (Lee, 2007), lightning formation (Mayer, 2001; Mayer, et al., 2005), and toilet tanks (Hegarty, et al., 2003; Mayer, et al., 2005). Other topics have included plant anatomy (Bartholome & Bromme, 2009), computer algorithms (Byrne, et al., 1999; Catrambone & Seay, 2002), foreign language (Plass, et al., 2003), and ATP synthesis (Munzer, et al., 2009).

Rather than sampling from this wide variety of topics, the approach in the present experiments was to select three physical systems with solid, movable components. Three topics rather than one were chosen to enable generalization to other similar physical systems. Physical systems with solid, movable components were chosen because they were amenable to diagrammatic representations of discrete system states.

***Background Questionnaire and Knowledge Pre-Test.*** Participants rated their knowledge on subjective rating scales (1-10) for each of the three topics. For each topic, there was a general-knowledge (valves, automobiles, and plumbing) and a specific-knowledge rating (dual-action air pumps, carburetors, and toilet tanks). An overall subjective knowledge score was computed by averaging the general and specific ratings. The questionnaire is shown in Appendix A.

Knowledge pre-tests for each topic were administered prior to the learning sessions. These consisted of four questions that were similar to those used after the learning session, including two questions that assessed knowledge of system structure and two questions that assessed knowledge of system behavior.

***Instructional conditions.*** Three instructional conditions were used. In the single diagram condition, participants viewed a single static diagram integrated with text (see Appendix B). Text segments were placed adjacent to the components and processes they described. Component names (e.g., "piston") that appeared within text segments were printed in red text, with red lines connecting the word to the component in the diagram. The same diagram appeared on each page of the instructions. In the multiphase diagrams condition, the system was depicted in different phases on each page of the instructions (see Appendix C).

The multiphase-plus-predictions condition contained identical pages as those used in the multiphase diagrams condition. The lesson was followed immediately by a series of prediction questions that were displayed on the screen adjacent to an image of the system. These prediction questions are similar to those that are used on typical transfer tests. Participants did not receive feedback after their responses. For the pump topic, the questions were as follows (correct answers are in italics):

- As the handle moves up, what direction does air flow in the lower conduit?
    - *To the left*

- Consider the following malfunction: What would happen if the piston rose too far and passed above the upper conduit? That is, what would happen to the pump's effectiveness?
  - *Air would move backwards through the system, coming from the discharge hose and back to the suction intake.*
  - *The inflated object would deflate.* (Both of these are acceptable answers).
- As the piston moves down, what direction does the air flow in the pump (clockwise or counterclockwise)?
  - *Clockwise*
- If air in the upper conduit were moving to the left, in what direction would air in the lower conduit move?
  - *To the right*

***Learning assessment.*** Two subject matter experts (SMEs) reviewed the lesson materials and the questions used for the post-test. They were asked to answer all of the questions after viewing the lessons. Questions that were answered incorrectly by both SMEs were discarded. The wording of several questions was modified in accordance with suggestions from the SMEs.

Performance on the post-test served as a proxy for measurement of learning. That is, post-test performance is related to – but does not directly measure – how much a given participant learned from the lessons. However, "learning" was used to refer to post-test performance.

Questions to assess learning were categorized *a priori* according to structure-behavior-function theory (Hegarty, et al., 2003; Heiser & Tversky, 2006; Liu & Hmelo-Silver, 2009). Appendix D shows the list of questions and their categorization for the pump topic. The criteria for categorization were as follows: Items that referred to motion were categorized as *behavioral,* whereas items that did not refer to motion were categorized as *structural*. Two experimenters independently categorized the test items. Categorizations were in agreement on 84% of the items, and the remaining items were resolved by discussion.

***Spatial ability tests.*** Three tests of visualization ability (Vz) were administered, including the Surface Development Test, the Paper Folding Test (Educational Testing Service, Ekstrom, et al., 1976), and the Mental Rotation Test (Peters, et al., 1995; Shepard & Metzler, 1971; Vandenberg & Kuse, 1978). Items from the Educational Testing Service tests are not displayed here due to copyright restrictions.

In the Surface Development test, participants must visualize how a two dimensional cardboard cutout would "develop" into a three dimensional object. Participants are asked to label edges in the three dimensional depiction with their corresponding labels shown on the two dimensional depiction.

The Paper Folding test (Ekstrom et al., 1976) has been frequently used in research on multimedia designs (e.g., Hegarty, et al., 2003; Hoffler & Leutner, 2010; Mayer & Sims, 1994). Participants see a depiction of a square piece of paper that has been folded along several lines and then hole-punched. Adjacent to this item are five unfolded pieces of paper with holes in various configurations. Participants mark the single item that would match the unfolded sample.

In the Mental Rotations Test (see participants see a depiction of a three-dimensional object composed of cubes that are connected in a linear fashion with 90° angles. Adjacent to the sample item is a test item; participants must determine whether the sample can be rotated three-dimensionally to match the test item.

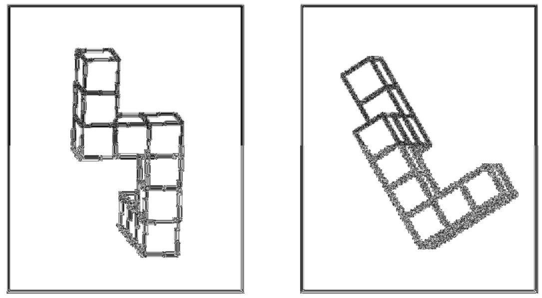

*Figure 4.* An example of test items from the Mental Rotations Test (Vandenberg & Kuse, 1978).

***Visuospatial working memory tests.*** The symmetry span test and the rotations span test were used as measures of visuospatial working memory (VSWM). Each test comprises two independent tasks. In the symmetry span test, participants are required to remember a spatial sequence of illuminated squares that appear in a 4x4 grid (see Figure 5). On each trial, 2-5 squares are successively illuminated. This is the primary task; scores on this span task serve as participants' span scores. The secondary task is a symmetry judgment. Symmetry judgments precede the illumination of each square in the sequence to be remembered. Participants are asked whether a grid of black and white squares is symmetric about its vertical axis (see *Figure 6*). After the entire sequence has been presented and symmetry judgments have been made, participants reproduce as much of the sequence as they can remember by clicking on the squares of a blank 4x4 grid. Participants receive one point for each item in the sequences that they are able to reproduce in full; they do not receive points for items that are not part of a complete correct sequence (i.e., there is no partial credit).

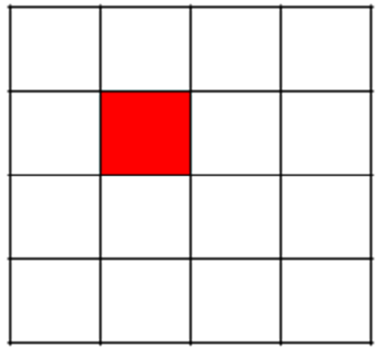

*Figure 5.* A sample stimulus for the span task that is performed during the symmetry span test. This depicts the illumination of a single square that would be followed by other squares in the sequence.

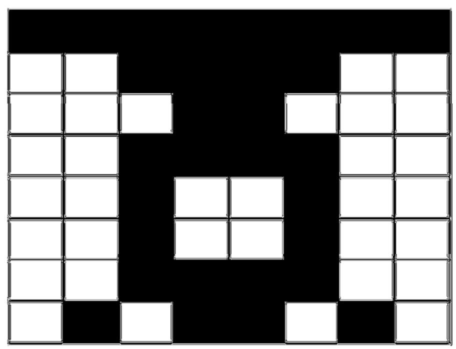

*Figure 6.* A sample stimulus for the symmetry judgment task that is performed during the symmetry span task.

The rotation span test uses a set of arrows for the storage task (see Figure 7). A sequence of arrows is displayed on a blank screen in a random order. At test, all eight of the arrows are shown, and participants click on arrowheads in the order in which they had appeared. For the processing task, which occurs before the presentation of each arrow, participants must determine whether letters would appear normal or mirrored when rotated to an upright position.

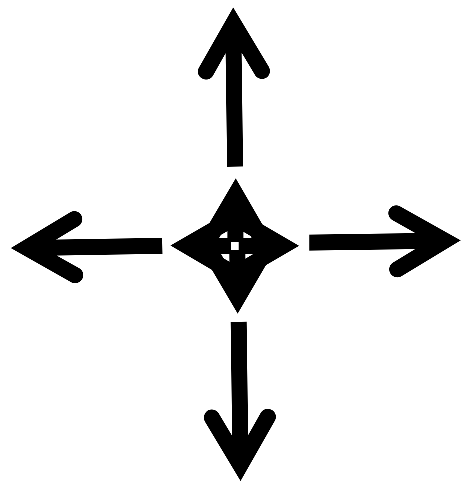

*Figure 7.* The set of arrows used in the rotation span test.

***Verbal ability tests.*** Tests of verbal ability included the Word Beginnings test, the Extended Range Vocabulary test (Ekstrom, et al., 1976), and the Cloze test (Taylor, 1957). The word beginnings test (Ekstrom, et al., 1976) assesses a participant's ability to rapidly think of as many words as he or she can that begin with a given three letters. The extended range vocabulary test (Ekstrom, et al., 1976) gives 48 sample words, each of which is accompanied by five other words. The participant is asked to mark the one of the five words that is synonymous with the sample word.

The Cloze test requires participants to fill in missing words in a text passage. Typically, every fifth or seventh word is replaced with a blank (e.g., Ackerman, Beier, & Bowen, 2000; Mikk, 2008). Proper nouns and numbers are not blanked; the preceding word is blanked instead. The first sentence in the passage is left unmodified.

Performance on the Cloze test is dependent on comprehension of the text passage, which is thought to depend on verbal fluency, specific knowledge of the passage topic, and relevant vocabulary (Ackerman, et al., 2000). Taylor (1957) found that performance on a Cloze test was substantially, positively correlated with performance on a standardized general aptitude test—the Armed Forces Qualifications Test, which was based on a number of indices, including work knowledge and arithmetical reasoning. Cloze performance was also correlated with multiple-choice performance on pre- and post-tests that assessed comprehension of the topic described in the Cloze passage.

For the present experiment, one Cloze test was derived from the opening passage of Oliver Twist (Dickens, 1867/2003). The passage contains 371 words, including the 47 words that were replaced by blank spaces (every seventh word). A second Cloze test was developed from the opening of the book, "So You Want to Start a Business?" (Hess & Goetz, 2008). The passage contained 368 words, including 46 blanks.

### 2.1.3 Design

A two-factor factorial design with incomplete blocks and Latin square assignment was used (Winer, 1971). Each participant learned about each of three topics and experienced each of three instructional conditions (see Table 1 for the assignment scheme).

Table 1. Instructional condition and topic assignments for participants 1-18 in Experiment 1.

| | | | Topic and Format Number | | |
|---|---|---|---|---|---|
| Participant | Block | Group Within Block | 1st Lesson | 2nd Lesson | 3rd Lesson |
| | | | | Block 1 | |
| 1 | 1 | 1 | Pump 1 | Carb. 2 | Toilet 3 |
| 2 | 1 | 2 | Pump 2 | Carb. 3 | Toilet 1 |
| 3 | 1 | 3 | Pump 3 | Carb. 1 | Toilet 2 |
| | | | | Block 2 | |
| 4 | 2 | 4 | Carb. 2 | Toilet 3 | Pump 1 |
| 5 | 2 | 5 | Carb. 3 | Toilet 1 | Pump 2 |
| 6 | 2 | 6 | Carb. 1 | Toilet 2 | Pump 3 |
| | | | | Block 3 | |
| 7 | 3 | 7 | Toilet 3 | Pump 1 | Carb. 2 |
| 8 | 3 | 8 | Toilet 1 | Pump 2 | Carb. 3 |
| 9 | 3 | 9 | Toilet 2 | Pump 3 | Carb. 1 |
| | | | | Block 4 | |
| 10 | 4 | 10 | Pump 1 | Carb. 3 | Toilet 2 |
| 11 | 4 | 11 | Pump 2 | Carb. 1 | Toilet 3 |
| 12 | 4 | 12 | Pump 3 | Carb. 2 | Toilet 1 |
| | | | | Block 5 | |
| 13 | 5 | 13 | Carb. 3 | Toilet 2 | Pump 1 |
| 14 | 5 | 14 | Carb. 1 | Toilet 3 | Pump 2 |
| 15 | 5 | 15 | Carb. 2 | Toilet 1 | Pump 3 |
| | | | | Block 6 | |
| 16 | 6 | 16 | Toilet 2 | Pump 1 | Carb. 3 |
| 17 | 6 | 17 | Toilet 3 | Pump 2 | Carb. 1 |
| 18 | 6 | 18 | Toilet 1 | Pump 3 | Carb. 2 |

*Note:* Each block is one of six orthogonal Latin squares. The numbers following each topic represent instructional formats, which were coded as 1, 2, and 3 for single diagram, multiphase diagram, and multiphase-plus-predictions, respectively. In Block 1, instructional format was assigned by Latin Square, while topic was held constant within columns. Blocks 2 and 3 are variants of Block 1, in which the columns were rotated by Latin square. Blocks 4-6 are variants of Blocks 1-3, respectively.

### 2.1.4 Procedure

The experiment was conducted in two sessions spaced seven days apart or as close as possible to seven days as possible (Minimum = 5, Maximum = 9). One to four students participated in any given session. In the first session, the following four items were administered: (a) a background questionnaire, (b) knowledge pre-test, (c) all measures of individual differences, (d) and the lessons for the three topics, including prediction questions for the applicable condition. The second session consisted of only the learning assessment.

At the beginning of the first session, participants were given a brief introduction to the experiment. The experimenter described the overall design of the two sessions. Next, participants completed the background questionnaire and answered the pre-tests questions. Next, cognitive ability tests were administered. To control for possible fatigue effects, the presentation order of tests for the three constructs (i.e., spatial ability, VSWM, and verbal ability) were counterbalanced across experiment session by Latin square. Finally, participants viewed the lesson materials on a computer. They used the keyboard to navigate between "pages" of the lesson. No time limit was imposed. The first session lasted approximately 3 hours.

After the prediction activity in the multiphase-plus-predictions condition, which immediately followed the corresponding lesson, participants were given sheets of paper with the printed diagram and prediction questions (one per page), and they were asked to indicate what strategy they used to answer the questions. At the top of the page, instructions read, "How did you create your answer for the question below? Please explain how you thought about the system as you created your answer."

The second session lasted approximately 45 minutes. Learning assessment questions were given on sheets of paper where participants recorded their answers. Questions for the three topics were given in the following order: pump, carburetor, toilet tank.

### 2.1.5 Scoring

All scores were standardized prior to statistical analysis. That is, each score was transformed by subtracting the mean and then dividing by the standard deviation.

***Learning assessment***. Participants were given zero points for incorrect answers, one point for fully correct answers, and 0.5 points for partially correct answers (some questions have two parts; see Appendix D). For the first 10 participants, answers were scored by two trained experimenters. Disagreements were resolved by discussion, and the grading rubric was modified accordingly. A single experimenter scored the responses from remaining participants.

***Retrospective reports of strategy.*** One experimenter scored the written responses. For each question, the number of written propositional statements was counted, and the number of drawn elements on the diagram was counted (e.g., arrows, lines, circles, etc…). These counts were summed across all of the prediction questions. A ratio of propositional statements to total responses was then calculated for each participant. The ratio was used to determine whether use of a visual strategy was associated with higher spatial ability or better performance on the learning assessment.

***Spatial ability tests***. Standard scoring procedures were followed for all tests. Participants received one point for each correct answer. For the Surface Development test, participants lost 0.2 points for each incorrect answer. For the Paper Folding test and

Mental Rotation Test, there was no penalty for incorrect answers. A composite spatial ability score was calculated by averaging the three standardized test scores.

***Verbal ability tests.*** Participants received one point for every correct answer on each of the three tests. In the Cloze test, answers were required to be grammatically correct and semantically acceptable in the given context (i.e., synonyms for the "correct" answer were acceptable). A composite verbal ability score was calculated by averaging the three standardized test scores.

***Visuospatial working memory tests.*** For each test, a participant's span score was calculated by the total number of list-memory items that were remembered correctly; items were counted as correct only if the entire series in which they appeared was reported correctly (Conway, et al., 2005). A participant's score on a VSWM test was discarded if his or her accuracy on the processing component was below 85% correct, because poor performance on the processing task might have been evidence that the participant was devoting too many cognitive resources to storage and rehearsal, while ignoring the processing task (Conway, et al., 2005). Scores for 10 participants were discarded from the Rotation Span task, and scores for 6 participants were discarded from the Symmetry Span task. Rotation Span data were lost for 4 participants due to a software malfunction. The composite span score was calculated by averaging the two test scores. None of the participants had scores discarded from both tests; thus, each participant had at least one span score from which to compute a composite VSWM score.

## 2.2 Results and Discussion

Cronbach's coefficient alpha was used to assess the reliability of the learning assessments for each of the three topics. Coefficient alpha provides an estimate of the lower bound of a test's reliability (Cortina, 1993). Cronbach's alpha coefficients of 0.7 or greater are generally accepted as an indication of adequate internal consistency (Nunnally & Bernstein, 1994). Alpha coefficients are shown in Table 2. For all topics, the internal consistency within the Structure questions and Behavior questions was below 0.7. A total score was also computed for each participant on each topic, based on the total percent correct for all questions. Reliabilities for these sets all exceeded 0.7. All of the data analyses described below were performed with scores from the separate structure and behavior sets, which had relatively low reliabilities. Analyses were also performed with the total score from each topic (not shown), but the pattern of results did not differ from the analyses of the separate structure and behavior sets.

Table 2. Cronbach's alpha coefficients for the learning assessment questions in Experiment 1.

| | α | | |
|---|---|---|---|
| Question Type | Pump | Carburetor | Toilet Tank |
| Structure | 0.62 | 0.58 | 0.54 |
| Behavior | 0.67 | 0.66 | 0.64 |
| Both (Total Score) | 0.77 | 0.76 | 0.71 |

### 2.2.1 Effects of Instructional Format, Topic, and Question Type

The plan for the analysis of variance was obtained from Winer (1971). A mixed effects generalized linear model was fit to the data. Factors included Format, Topic, Block (Lesson Order), Groups within Block, and Question Type. The six levels of Block corresponded to the six orthogonal Latin squares by which Topic and Format were assigned (see Table 1). Within each of these six Latin squares there are three rows; each row represents one of the 18 unique combinations of Topic x Format x Lesson Order. These rows were referred to as "Group within Blocks".

A main effect of Instructional Format was expected. Performance in the Multiphase-plus-predictions condition was expected to exceed performance in the Multiphase condition, which was expected to exceed performance in the Single phase condition. A main effect of Question Type was also expected, wherein performance on Structure questions would exceed performance on Behavior questions. The main effect of Instructional Format was expected to be qualified by an interaction with Question Type, wherein learning about system behavior but not structure is expected to be affected by the Format. Average scores and standard deviations are shown in Table 3 and Figure 8.

Table 3. Mean (SD) percent correct on the learning assessment questions from Experiment 1.

| | Instructional Format | | | |
|---|---|---|---|---|
| Topic | Single Diagram | Multiphase Diagrams | Multiphase & Prediction | Marginal Mean |
| | | Structure Questions | | |
| Pump | 41.67 (19.03) | 54.17 (23.79) | 56.77 (26.32) | 50.87 |
| Carburetor | 39.29 (23.92) | 37.50 (26.24) | 38.10 (27.52) | 38.30 |
| Toilet Tank | 38.69 (24.47) | 41.67 (21.86) | 42.86 (27.94) | 41.07 |
| *Marginal Mean* | 39.88 | 44.45 | 45.91 | 43.41 |
| | | Behavior Questions | | |
| Pump | 51.22 (17.91) | 54.51 (18.95) | 52.78 (24.72) | 52.84 |
| Carburetor | 35.81 (14.92) | 39.97 (17.92) | 49.09 (20.05) | 41.62 |
| Toilet Tank | 58.33 (23.17) | 55.09 (22.34) | 53.70 (24.88) | 55.71 |
| *Marginal Mean* | 48.45 | 49.86 | 51.86 | 50.06 |
| *Column Mean* | 44.17 | 47.15 | 48.88 | 46.73 |

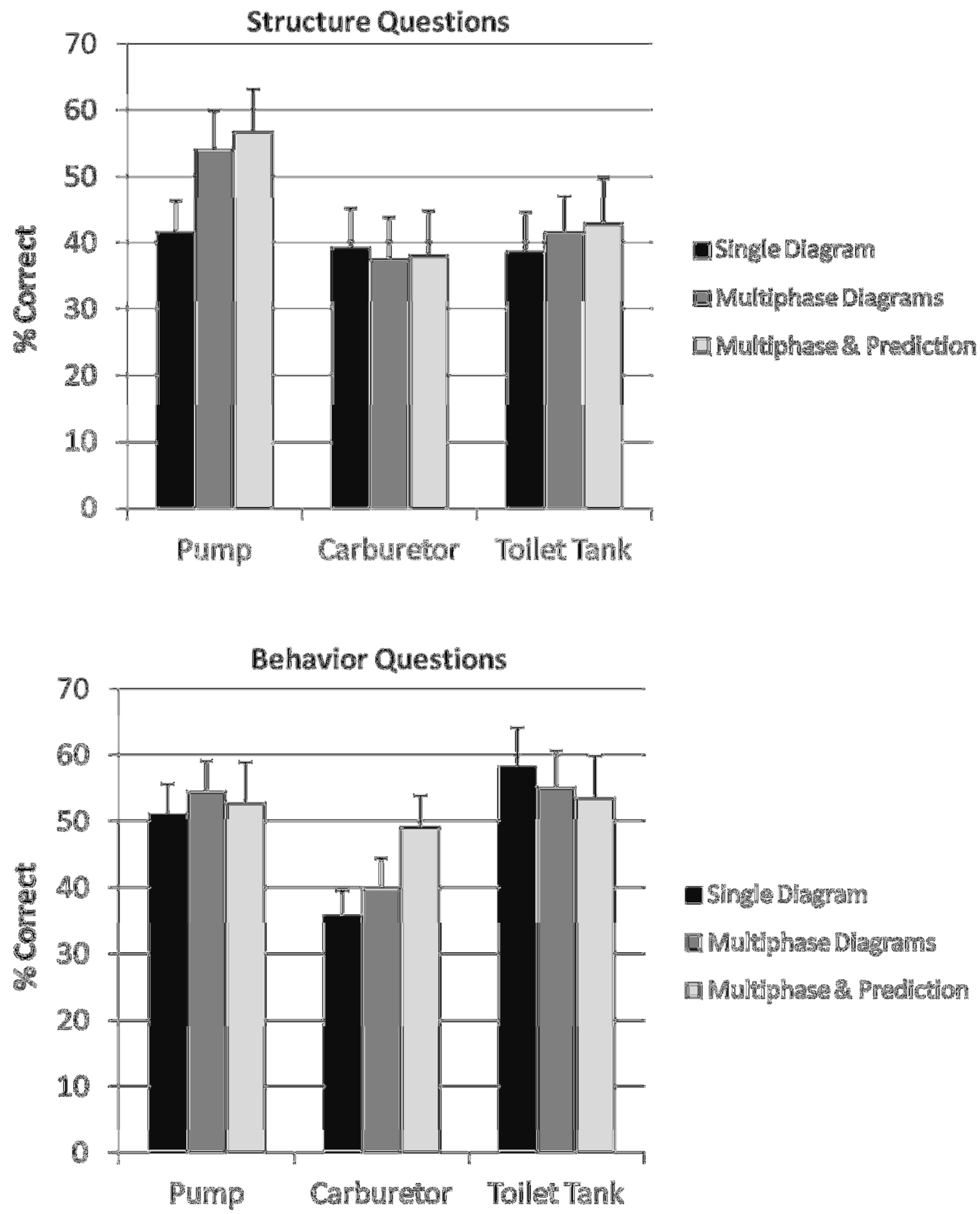


*Figure 8.* Performance on the learning assessment in Experiment 1 for Structure questions (upper panel) and Behavior questions (lower panel). Error bars represent 95% confidence intervals. Note that error bars for within-subject factors cannot be used to infer statistical significance (Cumming & Finch, 2005).

The effect of Format was significant, $\chi^2(2) = 7.67$, $p = .022$. A post-hoc t-test showed that participants performed more poorly in the single diagram condition than in the multiphase-plus-prediction conditions (see Table 4). Comparisons between other formats were not significant. Considering the Single Diagram condition as the baseline, the analyses showed that the addition of extra phase diagrams alone did not improve learning significantly. Instead, learning was improved when extra phase diagrams were supplemented with the prediction activity.

Table 4. Results for the post-hoc paired t-tests on Formats.

| Pair | *t* | df | *p* | Cohen's *d* | power |
|---|---|---|---|---|---|
| Single Diagram vs. Multiphase | 1.23 | 71 | 0.35 | 0.21 | .42 |
| Single Diagram vs. Multiphase&Prediction | 1.99 | 71 | 0.05 | 0.34 | .81 |
| Multiphase vs. Multiphase&Prediction | 0.55 | 71 | 0.58 | 0.09 | .12 |

*Note:* The critical alpha value with the Bonferroni correction for multiple comparisons is .0167

There was a significant effect of Question Type, $\chi^2(1) = 20.32$, $p < .001$. Contrary to predictions, participants performed better on Behavior questions than Structure questions. One possible reason for this effect is that the Structure questions always preceded the Behavior questions. Exposure to the Structure questions might have refreshed participants' memory of the system and assisted them later when they answered the Behavior questions. Another possibility is that many of the Structure questions required memory for the exact names of system components, whereas Behavior questions required knowledge of how the system worked. Participants might have forgotten the names of components, yet might still have been capable of reasoning about the system behavior. If participants were focused on learning how and why the systems worked, rather than on the details of their configurations and their component names, it is reasonable, in hindsight, to expect that they would perform better on the behavior questions relative to the structure questions.

There was a significant effect of Topic, $\chi^2(2) = 50.44$, $p < .001$. Paired t-tests on Topics indicated that the carburetor topic was significantly more difficult than the other two lessons (see Table 5). This was likely due to the fact that the carburetor was a more complex system (i.e., more element interactivity; Sweller & Chandler, 1994).

Table 5. Results for the post-hoc paired t-tests on the marginal means of Topic.

| Pair | *t* | df | *P* | Cohen's *d* |
|---|---|---|---|---|
| Pump vs. Carburetor | 5.30 | 71 | < .001 | 0.89 |
| Pump vs. Toilet Tank | 1.31 | 71 | 0.20 | 0.22 |
| Carburetor vs. Toilet Tank | 3.28 | 71 | 0.002 | 0.55 |

*Note:* The critical alpha value with the Bonferroni correction for multiple comparisons is .0167

### 2.2.2 Two-way Interactions

Significant interactions in the omnibus test were examined by analyses of simple main effects. The interaction between Topic and Question Type was significant, $\chi^2(2) = 16.30$, $p < .001$. The interaction was due to larger differences between Structure and Behavior questions in the toilet topic than in the other topics (see Table 6). Therefore, the main effect of Question Type was driven by greater differences within the toilet topic than within the other topics.

Table 6. Results of simple main effects analyses for the interaction between Topic and Question Type.

| Simple Main Effects | df (num) | df (den) | *F* | *P* | $\eta^2$ |
|---|---|---|---|---|---|
| Effect of Topic within Question Type | | | | | |
| Structure Questions | 2 | 142 | 7.44 | <.001 | .05 |
| Behavior Questions | 2 | 142 | 15.24 | <.001 | .08 |
| Effect of Question Type within Topic | | | | | |
| Pump | 1 | 71 | 0.70 | 0.88 | <.01 |
| Carburetor | 1 | 71 | 1.93 | 0.17 | <.01 |
| Toilet Tank | 1 | 71 | 23.10 | <.001 | .09 |

There was not a significant interaction between Format and Question Type, $\chi^2(2) = .96$, $p = .62$. The interaction between Format and Topic was significant, $\chi^2(4) = 11.59$, $p = .02$. The interaction was due to larger differences among topics within the Multiphase Diagram format than the other formats (see Table 7 and the middle group of data points in Figure 9). T-tests within the Multiphase format showed that performance was significantly worse on the carburetor topic than the pump topic, $t(142) = 3.63$, $p <.001$, and the toilet topic, $t(142) = 2.53$, $p = .012$. The difference between the pump and toilet topics was not significant, $t(142) = 1.05$, $p = .30$. Note that the interaction between Format and Topic was due to differences among Topics within one of the Formats, rather than differences among Formats within one of the Topics.

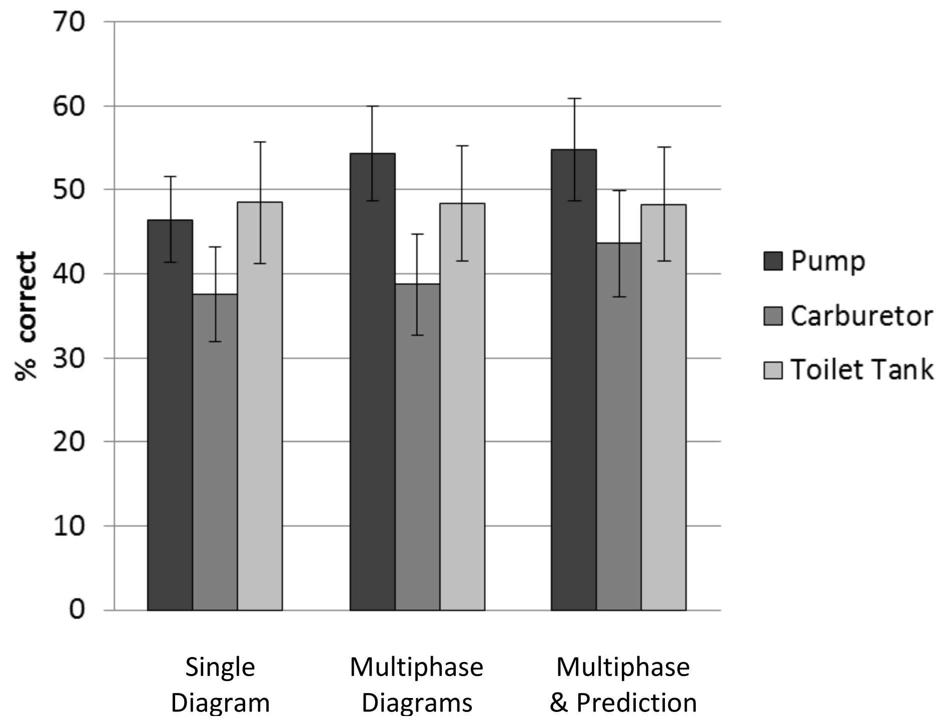


*Figure 9.* Interaction between Topic and Instructional Format.

Table 7. Results of simple main effects analyses for the interaction between Topic and Format.

| Simple Main Effects | df (num) | df (den) | *F* | *P* | $\eta^2$ |
|---|---|---|---|---|---|
| Effect of Format within Topic | | | | | |
| Pump | 2 | 69 | 1.39 | 0.25 | .04 |
| Carburetor | 2 | 69 | 0.62 | 0.54 | .02 |
| Toilet Tank | 2 | 69 | 0.01 | 0.99 | <.01 |
| Effect of Topic within Format | | | | | |
| Single Diagram | 2 | 69 | 2.52 | 0.09 | .07 |
| Multiphase Diagrams | 2 | 69 | 4.02 | 0.02 | .10 |
| Multiphase & Prediction | 2 | 69 | 1.52 | 0.23 | .04 |

### 2.2.3 Interaction among Topic, Format, and Question Type

There was a significant three-way interaction among Topic, Format, and Question Type, $\chi^2(4) = 12.65$, $p = .01$ (see Figure 8). This three-way interaction was examined by analyzing the interaction of Format by Question Type within each Topic. The interaction between Format and Question Type was not significant for the pump topic, $F(2,69) = 2.67$, $p = .08$, $\eta^2 = .02$, or for the toilet topic, $F(2,69) = 0.7$, $p = .49$, $\eta^2 = .01$, but the interaction was significant within the carburetor topic, $F(2,69) = 3.26$, $p = .04$, $\eta^2 = .02$. To examine this two-way interaction, a simple main effects analysis was performed on the data within the carburetor topic. For Structure questions, there was not a significant difference among Formats, $F(2,69) = .03$, $p = .97$, $\eta^2 < .01$, but for Behavior questions, there was a significant difference among Formats, $F(2,69) = 3.51$, $p = .035$, $\eta^2 = .09$. Paired t-test results (see Table 8) showed that performance on behavior questions was significantly better in the multiphase-plus-prediction condition than in the single diagram condition. Other comparisons were not significant. Thus, in accordance with predictions, performance on behavior questions in the carburetor topic was significantly better in the multiphase-plus-predictions condition than in the single phase condition. This finding mirrors the main effect of Format, and shows that the largest difference between single diagrams and multiphase-plus-predictions was for behavior questions on the carburetor topic.

However, as described above, the expected two-way interaction between Format and Question Type was not found (i.e., the two-way interaction was found only for the carburetor topic). There are various possible reasons why the interaction was found for the carburetor topic but not the other topics, such as differences among topics in the learning assessment items and complexities of the systems.

Table 8. Results for the paired t-tests on Instructional Formats within the Behavior questions on the Carburetor topic.

| Pair | *t* | df | *p* | Cohen's *d* |
|---|---|---|---|---|
| Single Diagram vs. Multiphase | 0.95 | 23 | 0.35 | .03 |
| Single Diagram vs. Multiphase&Prediction | 3.30 | 23 | 0.003 | .95 |
| Multiphase vs. Multiphase&Prediction | 1.78 | 23 | 0.09 | .51 |

*Note:* The critical alpha value with the Bonferroni correction for multiple comparisons is .0167.

### 2.2.4 Individual Differences: Descriptive Statistics and Correlations

SAT scores were reported by only 39 of the 72 participants, so the scores were not analyzed. Scores for each cognitive ability test were standardized prior to all analyses. Unstandardized means and standard deviations are shown in Table 9. Split-half reliability estimates for each test are shown in Table 10. Reliability of the Word Beginnings test was substantially lower than that of the other tests.

Table 9. Means and standard deviations for the measures of individual difference. Maximum possible scores are shown in parentheses.

| Test | Mean (Max) | SD |
|---|---|---|
| Cloze | 80.80 (100) | 11.61 |
| Extended Range Vocab. | 20.89 (48) | 7.60 |
| Word Beginnings | 28.24 (n/a) | 8.05 |
| Symmetry Span | 22.61 (60) | 9.32 |
| Rotation Span | 23.70 (60) | 8.76 |
| Mental Rotations | 11.06 (24) | 5.20 |
| Surface Development | 44.02 (60) | 13.83 |
| Paper Folding | 14.18 (20) | 3.31 |

*Note:* *Standard deviation was not reported.

Table 10. Split-half reliability estimates of individual differences tests in Experiment 1. The *r*-values represent correlations between the first and second half of each test.

| Test | *r* | Spearman-Brown Reliability Estimate |
|---|---|---|
| Cloze | 0.73 | 0.85 |
| Word Beginnings | 0.48 | 0.65 |
| Vocabulary | 0.73 | 0.84 |
| Paper Folding | 0.68 | 0.81 |
| Surface Development | 0.81 | 0.89 |
| Mental Rotations | 0.69 | 0.81 |

*Note.* The Spearman-Brown prophecy formula was used to estimate test reliability after correcting for total test length, which was halved prior to the computation of correlation.

Table 11 shows correlations between the standardized measures of individual differences, including verbal ability tests, spatial ability tests, visuospatial working memory tests, and each participant's grand mean for all of the tests. The Word Beginnings test showed poor convergent validity (i.e., its correlations with other verbal tests were low, and these correlations were only slightly higher than its correlation with the Surface Development test). Because the Word Beginnings test showed poor reliability (see Table 10) and poor convergent validity, it was excluded from further analyses.

Table 11. Correlations between standardized measures of individual differences in Experiment 1.

| | Vocab. | Cloze | Paper Folding | Surface Devel. | Mental Rotat. | Symm. Span | Rotation Span | **Item-Total** |
|---|---|---|---|---|---|---|---|---|
| Word Beginnings | 0.29* | 0.26* | 0.12 | 0.24* | -0.01 | 0.02 | -0.17 | **0.43**** |
| Vocabulary (ETS) | | 0.51** | 0.31** | 0.35** | 0.26* | 0.02 | -0.01 | **0.62**** |
| Cloze | | | 0.37** | 0.32** | 0.34** | -0.14 | -0.08 | **0.57**** |
| Paper Folding | | | | 0.54** | 0.45** | -0.01 | 0.10 | **0.59**** |
| Surface Development | | | | | 0.49** | 0.03 | 0.27* | **0.70**** |
| Mental Rotation | | | | | | -0.01 | 0.24 | **0.61**** |
| Symmetry Span | | | | | | | 0.36* | **0.30*** |
| Rotation Span | | | | | | | | **0.39**** |

*Note*: *p < .05. **p < .01

Composite scores for verbal ability, spatial ability, and VSWM were computed by averaging the standardized scores from the respective tests. Specifically, a Spatial Ability Composite was computed from standardized Surface Development, Paper Folding, and Mental Rotations scores. A Verbal Ability Composite was computed from standardized Extended Range Vocabulary and Cloze scores. VSWM was computed from standardized Symmetry Span and Rotation Span scores. Correlations between these composite scores are shown in Table 12.

Table 12. Correlations between Composite individual difference measures.

| | Spatial Ability | VSWM |
|---|---|---|
| Verbal Ability | .37** | -0.04 |
| Spatial Ability | | 0.06 |

*Note*: **p < .01

Participants' background knowledge of the topics was assessed with a knowledge pre-test (see Table 13) and self-report of knowledge on each topic (see Table 14). Reliabilities of the pre-tests were low: Cronbach's alpha coefficients for the pump, carburetor, and toilet topics were 0.21, 0.19, and 0.27, respectively. Correlations between these two measures for each topic are shown in Table 15. The low correlation for Pump items might indicate that one or both of the items were poor measures of the latent construct – namely, knowledge of the topic.

Table 13. Descriptive statistics for pre-test performance in Experiment 1.

| Topic | *M* (%) | *SD* |
|---|---|---|
| Pump | 14.93 | 12.96 |
| Carburetor | 9.03 | 15.22 |
| Toilet Tank | 20.49 | 23.35 |

Table 14. Descriptive statistics for subjective reports of knowledge on the topics in Experiment 1. Responses ranged from 1-10.

| Topic | *M* | *SD* |
|---|---|---|
| Pump | 2.38 | 1.52 |
| Carburetor | 2.53 | 1.88 |
| Toilet Tank | 3.72 | 2.42 |

Table 15. Correlations between pre-test scores and subjective knowledge ratings for the three topics in Experiment 1.

| Topic | *r* | *p* |
|---|---|---|
| Pump | 0.22 | 0.062 |
| Carburetor | 0.26 | 0.026 |
| Toilet Tank | 0.43 | <.001 |

### 2.2.5 Moderating Effect of Spatial Ability

It was predicted that spatial ability would be more predictive of performance on behavior than structure questions, because commonly used spatial ability tests require mental animation of images. This hypothesis would have been supported by an interaction between spatial ability and question type.

It was also expected that both spatial ability and VSWM would interact with instructional condition after partialling out variance associated with verbal ability. Spatial ability and VSWM were expected to be significantly more predictive of learning from multiphase diagrams – with or without predictions – than from the single diagram condition.

Effects of cognitive abilities on the experimental manipulations were evaluated by a hierarchical analysis of linear mixed effects models (Pinheiro & Bates, 2000). A linear mixed effects model accommodates both fixed and random effects. Experimental manipulations were treated as fixed effects and Participant was treated as a random effect.

Hierarchical analyses of fitted models were conducted separately on spatial ability and VSWM. The analysis for spatial ability assessed the effect of spatial ability while controlling for verbal ability. The analysis for VSWM assessed the effect of VSWM while controlling for verbal ability.

The order in which terms were removed from the full models was specified *a priori*. Results for the spatial ability hierarchical analysis are shown in Table 16. Verbal ability accounted for a significant amount of variance, and spatial ability accounted for a significant amount of variance after controlling for verbal ability. The expected interaction between spatial ability and question type was not obtained ($p$ = .78). Similarly, the expected interaction between spatial ability and instructional format was not significant ($p$ = .054). Although this interaction was not significant, the trend in the data was consistent with spatial ability enhancing the beneficial effect of the prediction activity (see Figure 10).

Table 16. Results for comparisons between hierarchical models including spatial ability and verbal ability.

| Model* | Terms in the Model | log Likelihood | Likelihood Ratio | *p* |
|---|---|---|---|---|
| 1 | SpatialAbility*Format*QuestionType | 89.94 | 2.87 | .24 |
| 2 | SpatialAbility*Format | 88.51 | 5.83 | .054 |
| 3 | SpatialAbility*QuestionType | 85.59 | 0.08 | .78 |
| 4 | SpatialAbility | 85.55 | 5.45 | .02 |
| 5 | VerbalAbility*Format*QuestionType | 82.83 | 0.69 | .71 |
| 6 | VerbalAbility*Format | 82.48 | 0.26 | .88 |
| 7 | VerbalAbility*QuestionType | 82.35 | 2.65 | .10 |
| 8 | VerbalAbility | 81.02 | 10.79 | .001 |
| 9 | Format*Topic*QuestionType | 75.63 | | |
| | Topic*QuestionType | | | |
| | Format*QuestionType | | | |
| | Format*Topic | | | |
| | BackgroundKnowledge | | | |
| | QuestionType | | | |
| | Topic | | | |
| | Format | | | |

*Note:* *Model #1 is the full model that includes all of the terms in the 2nd column. Each model includes the terms listed for that model, as well as all the terms below it. Likelihood ratios and p-values correspond to comparisons between the row on which they appear and the row beneath them.

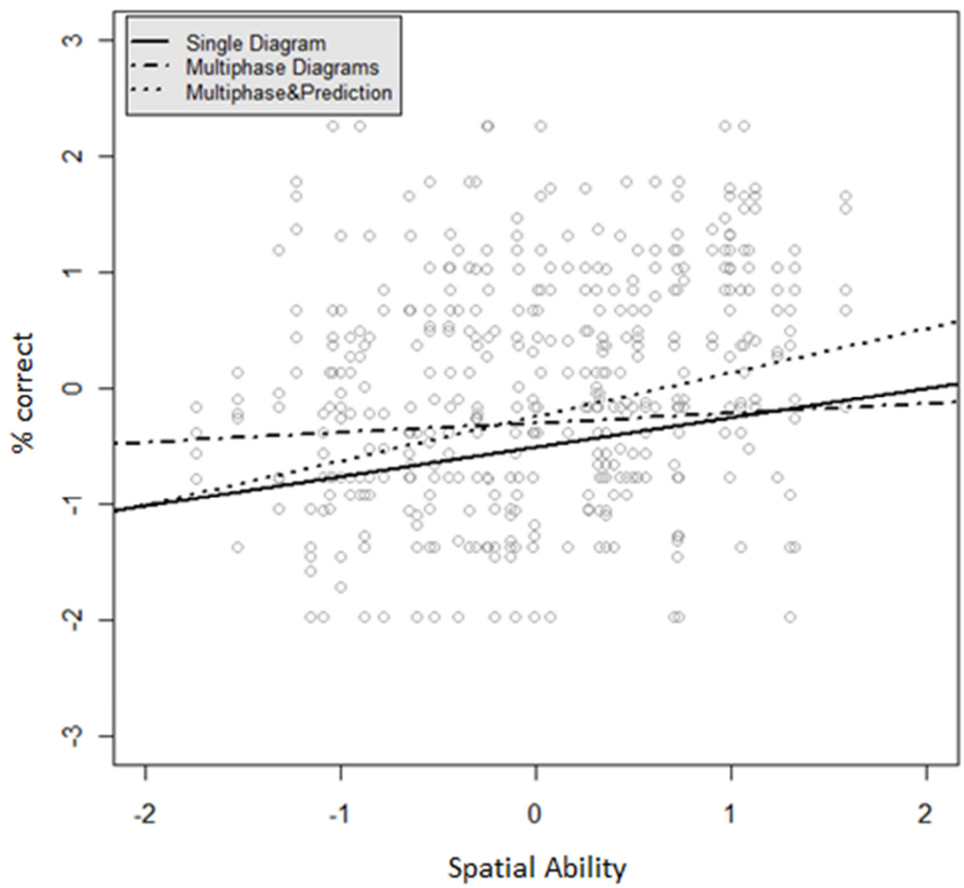


*Figure 10.* Effects of spatial ability and instructional format on post-test performance. Scores from the learning assessment and spatial ability tests were standardized.

It is possible that the effect of spatial ability on learning outcome in the multiphase-plus-prediction condition, which showed the highest correlation between spatial ability and learning (see Figure 10), was mediated by performance on the prediction questions. That is, high spatial ability might have caused better performance on the prediction questions, which in turn caused better performance on the tests of learning. Mediation requires significant correlations between each pair of terms in the mediation model (Baron & Kenny, 1986). This prerequisite was satisfied (see Figure 11).

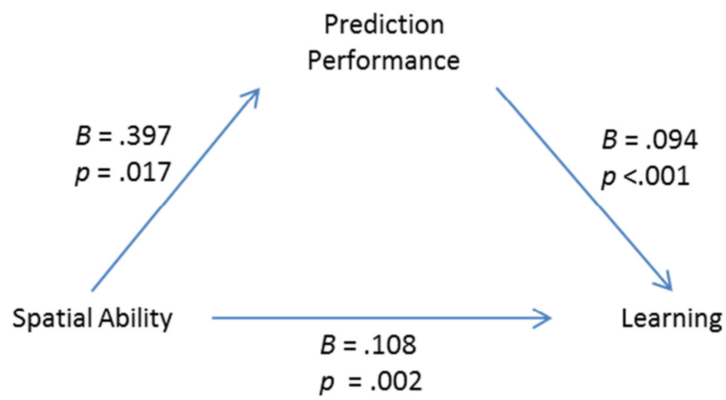


*Figure 11.* Model for spatial ability's effect on learning via mediation by performance on the prediction questions. *B*-weights and *p*-values refer to the results of three separate, pairwise regression analyses that were performed prior to the mediation analysis.

A mediation analysis was conducted with the "Mediation" package for R (Imai, Keele, & Yamamoto, 2010). The proportion of spatial ability's effect on learning that was mediated by prediction performance was 0.272, $CI_{95\%} = (0.025, 0.772)$. Complete results of the mediation analysis are shown in Table 17. The results indicate that, while a significant portion of spatial ability's total effect on learning was direct, another significant portion of its total effect on learning was mediated by performance in the prediction activity. This means that high prediction performance explains part of spatial ability's effect on learning in the multiphase-plus-predictions condition.

Table 17. Results of the mediation analysis. Prediction performance mediated a significant amount of spatial ability's effect on learning.

| | Coefficient | 95% CI | |
|---|---|---|---|
| Mediation by Prediction Performance | 0.030 | 0.003 | 0.066 |
| Direct effect of SA on Learning | 0.076 | 0.012 | 0.144 |
| Total effect of SA on Learning | 0.108 | 0.035 | 0.179 |
| Proportion of Total via Mediation | 0.272 | 0.025 | 0.772 |

*Note:* The "total effect" consists of the direct effect plus the mediated effect.

### 2.2.6 Moderating Effect of Visuospatial Working Memory

Results for the VSWM hierarchical analysis are shown in Table 18. After controlling for verbal ability, none of the VSWM terms accounted for significant variance in learning outcome. Moreover, the interaction between VSWM and Format was not significant ($p$ = .10; see Figure 12).

Table 18. Results for comparisons between hierarchical models including visuospatial working memory (VSWM) and verbal ability.

| Model* | Terms in the Model* | log Likelihood | Likelihood Ratio | *p* |
|---|---|---|---|---|
| 1 | VSWM*Format*QuestionType | 86.29 | 1.09 | 0.58 |
| 2 | VSWM*Format | 85.75 | 34.67 | 0.10 |
| 3 | VSWM*QuestionType | 83.41 | 1.17 | 0.28 |
| 4 | VSWM | 82.83 | < .001 | 0.98 |
| 5 | VerbalAbility*Format*QuestionType | 82.83 | | |
| | VerbalAbility*Format | | | |
| | VerbalAbility*QuestionType | | | |
| | VerbalAbility | | | |
| | Format*Topic*QuestionType | | | |
| | Topic*QuestionType | | | |
| | Format*QuestionType | | | |
| | Format*Topic | | | |
| | BackgroundKnowledge | | | |
| | QuestionType | | | |
| | Topic | | | |
| | Format | | | |

*Note:* *Model #1 is the full model that includes all of the terms in the 2nd column. Each model includes the terms listed for that model, as well as all the terms below it. Likelihood ratios and p-values correspond to comparisons between the row on which they appear and the row beneath them.

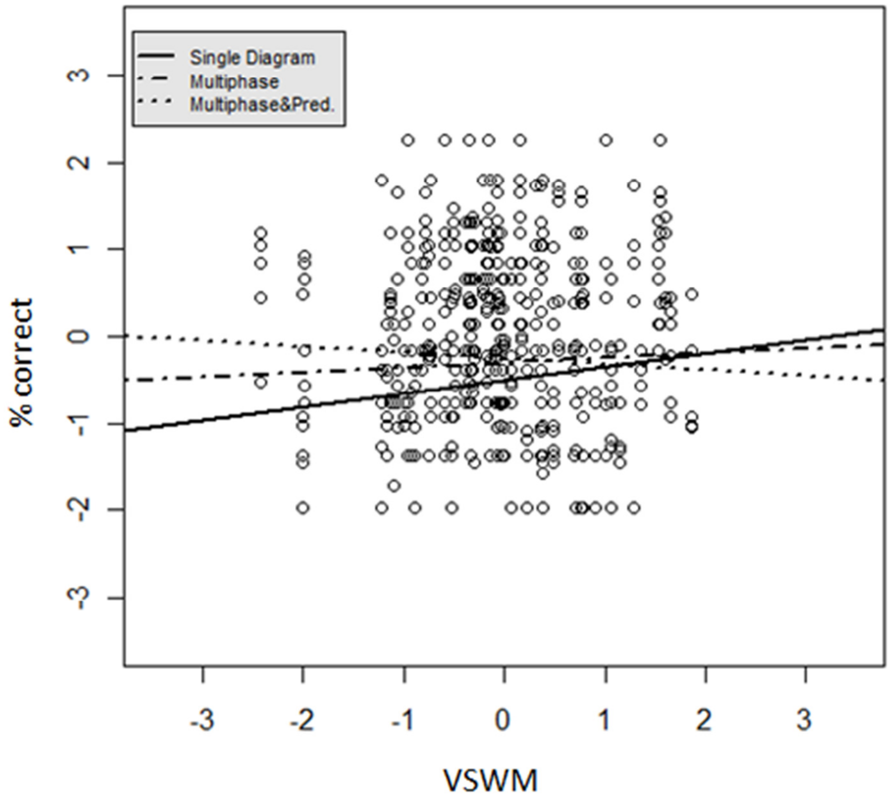


*Figure 12*. Non-significant ($p$ = .10) interaction between VSWM and instructional format in Experiment 1.

### 2.2.7 Strategy use during Predictions

After answering the prediction questions, participants were asked to explain how they had thought about the systems as they answered the questions. They returned to each question to write or draw an explanation of their thought process. The intent was to discover whether learners with higher spatial ability use a visual strategy, whereas learners with lower spatial ability used a verbal strategy.  Correlations with verbal ability and VSWM were also analyzed.

Many participants provided text-only explanations, while others provided a combination of text and drawings on the pictures (e.g., arrows representing movement and circles representing openings). For each participant, the number of propositional statements ($M$ = 3.83, $SD$ = 2.23) and number of drawn elements ($M$ = 3.58, $SD$ = 3.45) were counted; the sum of these was the total number of responses. The proportion of propositional statements was computed ($M$ = .63, $SD$ = .32) for each participant. The proportion did not correlate significantly with any of the individual difference measures (see Table 19). Thus, the data did not provide support for the hypothesis that learners with higher spatial ability would be more likely to rely on a visual strategy than a verbal strategy when answering the prediction questions.

Regardless of cognitive abilities, participants who showed evidence of using a visual strategy were more likely to answer the predictions correctly. That is, there was a significant correlation between prediction performance and the percentage of text statements, $t(62) = 3.62$, $r = -0.418$, $p < .001$. This finding is consistent with the hypothesis that participants who performed mental visualizations while answering the questions would show improved learning. Mental visualizations that were practiced during the prediction activity could have been useful during the subsequent post-test.

Table 19. Correlations between individual difference measures and percentage of propositional statements from participants' retrospective reports of strategy.

| Individual Difference | *r* | *p* |
|---|---|---|
| Verbal Ability | -0.04 | 0.73 |
| Spatial Ability | -0.21 | 0.09 |
| Working Memory | -0.15 | 0.23 |

## 2.3 Conclusion

Viewing multiphase diagrams and answering prediction questions were both expected to significantly improve learning, but results from Experiment 1 did not confirm these expectations. Instead, participants showed no significant benefit from multiphase diagrams over single phase diagrams, and no significant benefit from multiphase-plus-predictions over multiphase diagrams. Although the mean trended in the expected direction, the differences between that means were not statistically significant. Small effect sizes might have contributed to low statistical power (see Table 4). Therefore, it is possible that these manipulations could improve learning in a larger sample of participants.

The significant improvement in the multiphase-plus-predictions condition over the single diagram condition replicates the results of previous studies that have confounded the effects of phase diagrams and prediction activity (Hegarty, et al., 2003; Mayer & Gallini, 1990). The effect might be due to the small additive effects of each manipulation, or it might be due to an interaction between the two manipulations. The present study extends results of previous studies (Hegarty, et al., 2003; Mayer & Gallini, 1990) by showing that predictions or multiphase diagrams alone might not yield a large benefit over single phase diagrams.

Significant correlations were found among performance on prediction questions, post-test performance, and spatial ability. As expected, spatial ability was predictive of post-test performance even after controlling for verbal ability. The effect of spatial ability on post-test performance was partially mediated by performance on the prediction activity. The significant partial mediation suggests that spatial ability improved performance on the prediction questions, which in turn improved performance in the learning assessment. Participants with higher spatial ability might have been better able to perform mental simulations of motion during the prediction activity; this might have led to a better mental model of the system, which in turn led to better performance on the post-test period.

Although participants with higher spatial ability performed better on the prediction questions, they did not appear to use a different strategy than participants with lower spatial ability to answer those prediction questions. That is, there were not significant correlations between retrospective strategy reports and individual differences in spatial or verbal ability. The trend towards a negative correlation ($p$ = .09) between spatial ability and the number of text statements (as a percentage of text statements plus drawn elements) was in the opposite direction of that found by Hegarty and Steinhoff (1997), who reported that learners with lower spatial ability were more likely to draw on pulley diagrams while reasoning about the pulleys. One critical difference between the present study and the Hegarty and Steinhoff study is the participants drew *after* rather than *during* the reasoning process, respectively. When reasoning about the pulley system, participants with lower spatial ability might have benefited from the external memory aids embodied in the arrows that they drew.

Participants in the Hegarty and Steinhoff (1997) study were reasoning about familiar pulley systems, rather than learning about unfamiliar systems such as carburetors. If reasoning about familiar systems by low-spatial participants was improved by arrows, then learning about unfamiliar systems might also be improved by arrows. Specifically, learners with lower spatial ability might perform mental animations more accurately with arrows than without arrows. The arrows might provide processing cues that facilitate mental animations and improve understanding of causality. One condition in Experiment 2 included depictive arrows, with the expectation that they would benefit learners with lower spatial ability more than learners with higher spatial ability. Additionally, participants in two conditions in Experiment 2 were explicitly prompted to attempt mental simulations of motion while viewing the lesson materials.

# CHAPTER 3: EXPERIMENT 2

Participants' spatial abilities might affect their ability to accurately generate mental animations while answering prediction questions. If the animations are inaccurate, then participants' mental models would not improve from the prediction activity. In Experiment 1, this expectation was supported by the finding that prediction performance partially mediated the effect of spatial ability on learning outcome.

Mental animations might be generated more accurately if they are performed while viewing the lesson materials rather than afterwards (i.e., during the prediction activity). The mental animation process might be facilitated by the presence of the text describing the system (Gyselinck & Tardieu, 1999) and the presence of multiphase diagrams that can be compared; neither of these elements were presented during the prediction activity in Experiment 1. Furthermore, mental animations might be generated more accurately if arrows are included in the diagrams to depict the correct direction of motion.

In Experiment 1, participants whose retrospective reports suggested that they visualized movement were significantly more likely to answer the prediction questions correctly. In Experiment 2, participants were encouraged in some conditions to visualize movement in the system while viewing the lessons, prior to answering the prediction questions. If participants complied with instructions to visualize motion, they would be expected to perform better on the subsequent prediction questions as well as the delayed learning assessment. By generating mental animations, participants would be constructing a mental model (Hegarty & Just, 1993). Later during the post-test, simulations could be run on the mental model to determine how the system would behave in novel conditions.

Experiment 2 was designed to determine whether explicit directions to generate mental animations would improve learning. Additionally, it was designed to determine whether the inclusion of arrows in diagrams would assist learners, especially those with lower spatial ability. Arrows depict movement, and might therefore reduce the difficulty and errors associated with the generation of mental animations. Arrows might be particularly beneficial for learners with lower ability who would otherwise struggle to perform the mental animations accurately. Thus, an interaction was expected between spatial ability and instructional format, wherein spatial ability would be more predictive of learning from diagrams without arrows than with arrows.

## 3.1 Methods

Wherever applicable, methods for Experiment 2 were the same as those in Experiment 1 (i.e., same participant pool, background questionnaire, pre-test items, post-test items, ability tests, and procedure).

### 3.1.1 Participants

Seventy-two undergraduate students from the Georgia Institute of Technology participated in the experiment (M = 47, F = 23, Unreported = 2). They were compensated with extra credit for the psychology courses in which they were enrolled. The five participants who performed the best on the tests received $10 each.

### 3.1.2 Materials, Procedure, and Design

Participants learned about pumps, carburetors, and toilet tanks from one of three instructional conditions. A background questionnaire and knowledge pre-test for each topic were administered (see Appendix A). The tests of cognitive abilities that were administered in Experiment 1 were also administered in Experiment 2.

In each of the three instructional conditions, multiphase diagrams were shown, and participants were asked to answer predictions questions after the lessons. Table 20 lists the attributes of the experimental conditions. The multiphase condition consisted of multiphase diagrams and predictions (a replication of the third condition from Experiment 1, see Appendix C). In the second and third conditions, participants were explicitly prompted to visualize motion in the system with the following directions: "As you view the following lessons, please try to visualize motion of the system as it is described in the text or indicated in the diagram. This might help you learn the material better and perform better on the tests." In the third condition, arrows were added to the diagrams to depict motion (see Appendix E).

Table 20. Instructional conditions for Experiment 2.

| Condition Name | Description |
|---|---|
| Multiphase | Multiphase diagrams + predictions |
| Simulation | Multiphase diagrams + predictions + mental animations |
| Arrows | Multiphase diagrams + predictions + mental animations + arrows |

To prevent participants from inappropriately carrying over the instructions to visualize motion from one condition to the next, the multiphase condition was always presented first. The other two conditions were presented in counterbalanced order. Lesson topic was also counterbalanced (see Table 21). It is possible that some participants may have spontaneously visualized motion when answering prediction questions in the multiphase condition. If this were to occur, equivalent learning outcomes would be expected from the multiphase and simulation conditions for that participant.

Table 21. Condition assignment and orders for the first 12 participants in Experiment 2. Multiphase, Simulation, and Arrows Formats were coded as 1, 2, and 3, respectively.

| | | | Topic and Format Number | | |
|---|---|---|---|---|---|
| Participant | Block | Group within Block | 1st Lesson | 2nd Lesson | 3rd Lesson |
| 1 | 1 | 1 | Pump 1 | Carb. 2 | Toilet 3 |
| 2 | 1 | 2 | Pump 1 | Toilet 3 | Carb. 2 |
| 3 | 2 | 3 | Carb. 1 | Pump 2 | Toilet 3 |
| 4 | 2 | 4 | Carb. 1 | Toilet 3 | Pump 2 |
| 5 | 3 | 5 | Toilet 1 | Pump 2 | Carb 3 |
| 6 | 3 | 6 | Toilet 1 | Carb 3 | Pump 2 |
| 7 | 4 | 7 | Pump 1 | Toilet 2 | Carb. 3 |
| 8 | 4 | 8 | Pump 1 | Carb. 3 | Toilet 2 |
| 9 | 5 | 9 | Carb. 1 | Toilet 2 | Pump 3 |
| 10 | 5 | 10 | Carb. 1 | Pump 3 | Toilet 2 |
| 11 | 6 | 11 | Toilet 1 | Carb. 2 | Pump 3 |
| 12 | 6 | 12 | Toilet 1 | Pump 3 | Carb. 2 |

*Note:* The numbers following each topic represent instructional formats, which were coded as 1, 2, and 3 for multiphase, animation, and arrows, respectively. Participants within each Block received the same combination of topics and formats. The orders of the second and third lessons within each Block were counterbalanced by Group (within Block). This assignment scheme was repeated for the remaining 60 participants.

In all three conditions, participants answered prediction questions after viewing the lesson materials (see Appendix D). Each prediction question was presented with a single phase diagram of the system.

Rather than repeating the retrospective report of visual/verbal strategy use that was implemented in Experiment 1, participants in Experiment 2 were asked to rate on a scale from 1 to 10 how much they tried to mentally simulate motion in the diagrams. The following phrasing was used: “To what extent did you attempt to visualize motion in the diagrams?” Response options ranged from “very little” (1) to “very much” (10).

The learning assessment (i.e., post-test performance measure), experiment procedures, and scoring methods were the same as those reported above in Experiment 1.

### 3.2 Results and Discussion

Internal consistencies for the learning assessment items are shown in Table 22. They are similar to those observed in Experiment 1. For each topic, the internal consistency within the Structure questions and Behavior questions was below 0.7, which might indicate poor internal consistency (Nunnally & Bernstein, 1994). A total score was also computed for each participant on each topic, based on the total percent correct for all questions. Reliabilities for these sets were much closer to 0.7 (0.67 or greater).

All of the data analyses described below were performed with scores from the separate structure and behavior sets, which had relatively low reliabilities. Analyses were also performed with the total score from each topic (not shown), but the pattern of results did not differ from the analyses of the separate structure and behavior sets.

Table 22. Internal consistency for the learning assessment in Experiment 2.

| | α | | |
|---|---|---|---|
| Question Set | Pump | Carburetor | Toilet Tank |
| Structure | 0.59 | 0.45 | 0.58 |
| Behavior | 0.56 | 0.64 | 0.57 |
| Both (Total Score) | 0.72 | 0.73 | 0.67 |

### 3.2.1 Subjective Ratings of Attempts to Visualize Motion

Immediately after each lesson, participants rated the extent to which they attempted to visualize motion in the diagrams while viewing the lessons. This was a manipulation check to determine whether the explicit prompts to visualize motion were effective in evoking different levels of mental simulations in the conditions. Before the first lesson, which was always the multiphase condition, participants had not been prompted to visualize motion. Before the second and third lessons, they were prompted to visualize motion. Therefore, ratings were expected to be lower in the first lesson than in the second and third lessons, and ratings were not expected to differ between the second and third lessons. Means and standard deviations of the ratings are shown in Table 23.

There was not a significant difference in ratings of attempts to visualize motion among Formats, $F(2,142) = 2.51$, $p = .085$, $\eta^2 = .01$, $power = .17$. Given this null result, it is unclear whether the instructional conditions (i.e., Formats) had any effect on levels of mental visualizations. The manipulation might have been entirely ineffective. Alternatively it is possible that the ratings were simply invalid or unreliable. However, this interpretation is unlikely, given the significant correlation between ratings and learning outcome, $F(1,214) = 35.73$, $p < .001$, $R^2 = .145$. This correlation suggests that attempts to visualize motion were indeed effective in improving mental model formation.

Table 23. Subjective ratings of extent to which participants attempted to visualize motion in diagrams.

| Instructional Format | *M* | *SD* |
|---|---|---|
| Multiphase | 7.59 | 2.15 |
| Simulation | 8.37 | 1.55 |
| Arrows | 8.56 | 1.63 |

Spatial ability was not correlated with ratings, $t(214) = 1.98$, $p = .16$, $R^2 = .009$, nor was VSWM, $t(214) = .46$, $p = .49$, $R^2 = .002$. Thus, the data did not suggest that learners with higher spatial ability were more likely that those with lower spatial ability to use their ability for mental visualization.

### 3.2.2 Effects of Instructional Format, Topic, and Question Type

Data were analyzed with a generalized linear model fitted with Type 1 Sum of Squares. Participant was nested within Group within Block (see Table 21). Instructional Format, Topic, and Question Type were all treated as within-subjects factors. Data are displayed in Figure 13 and Table 24.

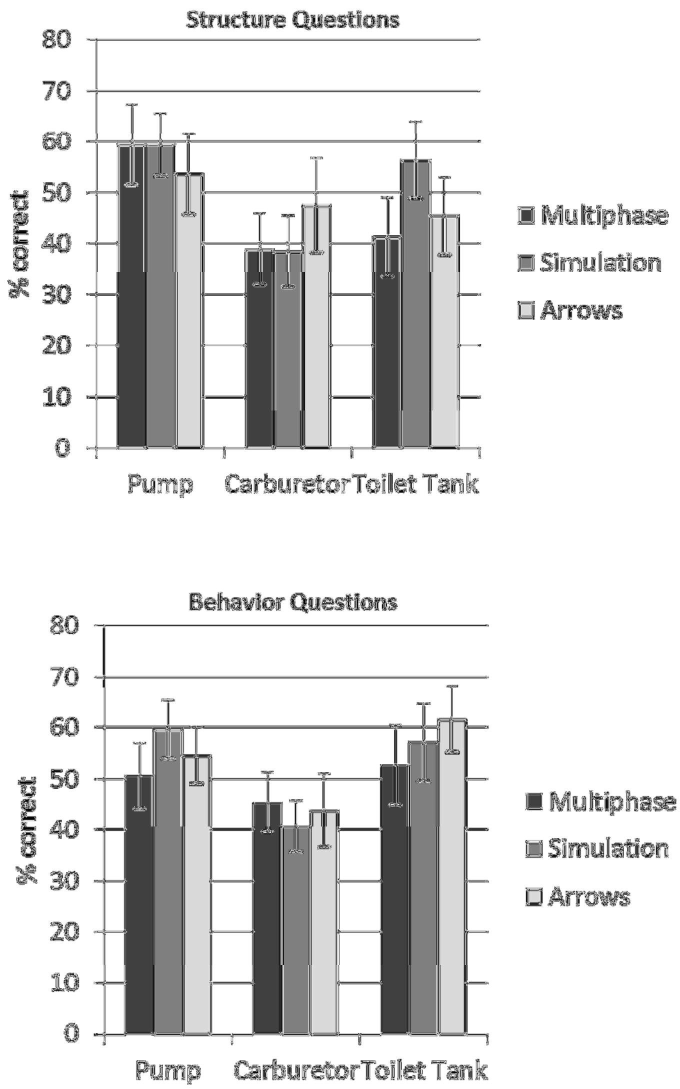


*Figure 13.* Performance on the learning assessment in Experiment 2 for Structure questions (upper panel) and Behavior questions (lower panel). Error bars represent 95% confidence intervals.

Table 24. Mean (SD) percent correct on the learning assessment questions from Experiment 2.

| | Instructional Format | | | |
|---|---|---|---|---|
| Topic | Multiphase Diagrams | Simulation | Arrows | Marginal Mean |
| | Structure Questions | | | |
| Pump | 59.38 (24.52) | 59.38 (19.24) | 53.65 (24.58) | 57.47 |
| Carburetor | 38.99 (22.19) | 38.57 (24.36) | 47.62 (24.98) | 40.97 |
| Toilet Tank | 41.37 (24.38) | 56.35 (20.48) | 45.48 (26.63) | 46.83 |
| *Marginal Mean* | 46.58 | 49.95 | 48.73 | 48.42 |
| | Behavior Questions | | | |
| Pump | 50.69 (19.99) | 59.55 (17.91) | 54.51 (17.15) | 54.92 |
| Carburetor | 45.44 (18.08) | 40.73 (17.42) | 43.75 (19.53) | 43.06 |
| Toilet Tank | 52.78 (24.35) | 57.1 (20.59) | 61.67 (22.57) | 57.56 |
| *Marginal Mean* | 49.67 | 51.09 | 54.80 | 51.85 |
| *Column Mean* | 48.11 | 50.52 | 51.77 | 50.13 |

The effect of Format was not significant, $\chi^2(2) = 5.20$, $p = .074$. It is possible that depictive arrows and prompts to simulate motion were ineffective manipulations. It is also possible that the present experiment lacked sufficient power to detect a small effect size.

As in Experiment 1, participants performed better on questions of system behavior than structure, $\chi^2(1) = 6.60$, $p = .01$. There was also a significant effect of Topic, $\chi^2(2) = 79.73$, $p < .001$. Paired t-tests showed that the participants performed more poorly on the carburetor topic than the other topics (see Table 25). The difference between the pump and toilet topics was not significant at the adjusted alpha level (.0167).

Table 25. Results for the post-hoc paired t-tests on the marginal means of Topic.

| Pair | *t* | df | *p* | Cohen's *d* |
|---|---|---|---|---|
| Pump vs. Carburetor | 8.15 | 71 | < .001 | 1.37 |
| Pump vs. Toilet Tank | 2.12 | 71 | 0.037 | 0.36 |
| Carburetor vs. Toilet Tank | 4.76 | 71 | < .001 | 0.79 |

*Note:* The critical alpha value with the Bonferroni correction for multiple comparisons is .0167

### 3.2.3 Interactions among Format, Topic, and Question Type

The interaction between Topic and Format was not significant, $\chi^2(4) = 3.49$, $p = .48$, but the interaction between Topic and Question Type was significant, $\chi^2(2) = 17.06$, $p < .001$. Analyses of simple main effects showed that the interaction was due to larger differences between Question Types for the toilet topic than the other topics (see Table 26). Within the toilet topic, performance was better on Behavior questions than Structure questions. This replicates the effect from Experiment 1, and shows that the main effect of Question Type was primarily driven by differences within the toilet topic.

Table 26. Results for simple main effects analysis of the interaction between Topic and Question Type.

| Simple Main Effects | df (num) | df (den) | *F* | *P* | $\eta^2$ |
|---|---|---|---|---|---|
| Effect of Topic within Question Type | | | | | |
| Structure Questions | 2 | 142 | 17.98 | <.001 | .077 |
| Behavior Questions | 2 | 142 | 18.95 | <.001 | .093 |
| Effect of Question Type within Topic | | | | | |
| Pump | 1 | 71 | 1.16 | 0.29 | .004 |
| Carburetor | 1 | 71 | 0.81 | 0.37 | .002 |
| Toilet Tank | 1 | 71 | 11.28 | 0.001 | .049 |

The interaction between Format and Question Type was not significant, $\chi^2(2) = 1.21$, $p = .55$. Likewise, the three-way interaction between Format, Topic, and Question Type was not significant, $\chi^2(4) = 5.01$, $p = .29$.

### 3.2.4 Individual Differences: Descriptive Statistics and Correlations

Unstandardized means and standard deviations are shown in Table 27. All scores were standardized prior to statistical analyses. Four participants did not attempt the second half of the Surface Development test, so their scores for the entire test were discarded. Split-half reliability estimates for each test are shown in Table 28.

Table 27. Means and standard deviations of the measures of individual differences in Experiment 2.

| Test | Mean (Max) | SD |
|---|---|---|
| Cloze | 79.79 (100) | 13.58 |
| Extended Range Vocabulary | 24.87 (48) | 12.63 |
| Word Beginnings | 25.15 (n/a) | 7.63 |
| Symmetry Span | 23.24 (60) | 8.85 |
| Rotation Span | 23.00 (60) | 8.93 |
| Mental Rotations | 12.60 (24) | 4.76 |
| Surface Development | 45.70 (60) | 13.47 |
| Paper Folding | 13.88 (20) | 3.42 |

Table 28. Split-half reliability estimates of individual differences tests in Experiment 2. The r-values represent correlations between the first and second half of each test.

| Test | *r* | Spearman-Brown Reliability Estimate |
|---|---|---|
| Cloze | 0.68 | 0.81 |
| Word Beginnings | 0.49 | 0.66 |
| Vocabulary | 0.71 | 0.83 |
| Paper Folding | 0.78 | 0.88 |
| Surface Development | 0.74 | 0.85 |
| Mental Rotations | 0.55 | 0.71 |

As in Experiment 1, the Word Beginnings test had low reliability. In this Experiment, though, the Mental Rotations test also had low reliability (see Table 28). Correlations between individual tests are shown in Table 29, and correlations between composite factors are shown in Table 30. The Spatial Ability Composite was the average of the standardized Paper Folding, Surface Development, and Mental Rotations tests. Composite scores for the four participants who did not complete the Surface Development Test were computed from the other two tests. The Verbal Ability Composite was the average of the standardized Word Beginnings, Extended Range Vocabulary, and Cloze tests. The visuospatial working memory (VSWM) composite was the average of the standardized Symmetry Span and Rotation Span tests.

Table 29. Correlations between standardized measures of individual differences in Experiment 2.

| | Vocab. | Cloze | Paper Folding | Surface Devel. | Mental Rotat. | Symm. Span | Rotation Span | **Item-Total** |
|---|---|---|---|---|---|---|---|---|
| Word Beginnings | 0.7** | 0.52** | 0.2 | 0.17 | 0.12 | -0.04 | -0.26* | 0.47** |
| Vocabulary (ETS) | | 0.53** | -0.04 | 0.06 | 0.07 | -0.08 | -0.13 | 0.46** |
| Cloze | | | -0.03 | -0.03 | -0.01 | -0.19 | -0.28* | 0.35** |
| Paper Folding | | | | 0.39** | 0.28* | 0.17 | -0.06 | 0.48** |
| Surface Development | | | | | 0.51** | 0.35** | 0.02 | 0.68** |
| Mental Rotation | | | | | | 0.19 | -0.05 | 0.59** |
| Symmetry Span | | | | | | | 0.45** | 0.52** |
| Rotation Span | | | | | | | | 0.28* |

*Note*: *p < *.05*. **p < *.01*

Table 30. Correlations between composite measures of individual differences in Experiment 2.

| | Spatial Ability | VSWM |
|---|---|---|
| Verbal Ability | 0.51** | 0.08 |
| Spatial Ability | | 0.13 |

*Note*: **p < .01

Background knowledge of each topic was assessed with a pre-test (see Table 31) and a subjective self-report (see Table 32). Cronbach's alpha coefficients for the pump, carburetor, and toilet knowledge pre-tests were 0.46, 0.52, and 0.50, respectively. Correlations between pre-test performance and subjective knowledge ratings for each topic are shown in Table 33. The low correlation for Pump items, which replicates the finding in Experiment 1, might indicate that one or both of the items were poor measures of background knowledge about dual action pumps.

Table 31. Descriptive statistics for pre-test performance in Experiment 2.

| Topic | *M* (%) | *SD* |
|---|---|---|
| Pump | 16.32 | 12.69 |
| Carburetor | 7.99 | 14.43 |
| Toilet Tank | 20.49 | 25.64 |

Table 32. Descriptive statistics for subjective reports of knowledge on the topics in Experiment 2. Responses ranged from 1-10.

| Topic | *M* | *SD* |
|---|---|---|
| Pump | 2.06 | 1.75 |
| Carburetor | 2.39 | 2.19 |
| Toilet Tank | 3.65 | 2.54 |

Table 33. Correlations between objective and subjective measures of background knowledge on the three topics in Experiment 2.

| Topic | *r* | *p* |
|---|---|---|
| Pump | 0.22 | 0.11 |
| Carburetor | 0.40 | <.001 |
| Toilet Tank | 0.53 | <.001 |

### 3.2.5 Individual Differences: Moderating Effects

The analyses performed in Experiment 1 were repeated in Experiment 2. Hierarchical linear mixed effects models were fit to the data. The first analysis examined the effects of spatial ability and verbal ability (see Table 34). Verbal ability accounted for a significant amount of variance in learning outcome, but did not interact with Format or Question Type.

While controlling for verbal ability, spatial ability accounted for significant amount of variance in learning outcome. Contrary to expectations, spatial ability did not interact significantly with Format or Question Type (see Table 34 and Figure 14).

Table 34. Results for comparisons between hierarchical models including spatial ability and verbal ability in Experiment 2.

| Model | Terms in the Model | log Likelihood | Likelihood Ratio | *p* |
|---|---|---|---|---|
| 1 | SpatialAbility*Format*QuestionType | 141.10 | 1.71 | 0.42 |
| 2 | SpatialAbility*Format | 140.25 | 4.65 | 0.10 |
| 3 | SpatialAbility*QuestionType | 137.93 | 2.83 | 0.09 |
| 4 | SpatialAbility | 136.51 | 15.59 | <.001 |
| 5 | VerbalAbility*Format*QuestionType | 128.71 | 1.79 | 0.41 |
| 6 | VerbalAbility*Format | 127.82 | 0.27 | 0.87 |
| 7 | VerbalAbility*QuestionType | 127.68 | 0.45 | 0.50 |
| 8 | VerbalAbility | 127.46 | 11.01 | 0.001 |
| 9 | Format*Topic*QuestionType | 121.95 | | |
| | Topic*QuestionType | | | |
| | Format*QuestionType | | | |
| | Format*Topic | | | |
| | BackgroundKnowledge | | | |
| | QuestionType | | | |
| | Topic | | | |
| | Format | | | |

*Note:* Model #1 is the full model that includes all of the terms in the 2nd column. Each model includes the terms listed for that model, as well as all the terms below it. Likelihood ratios and p-values correspond to comparisons between the row on which they appear and the row beneath them.

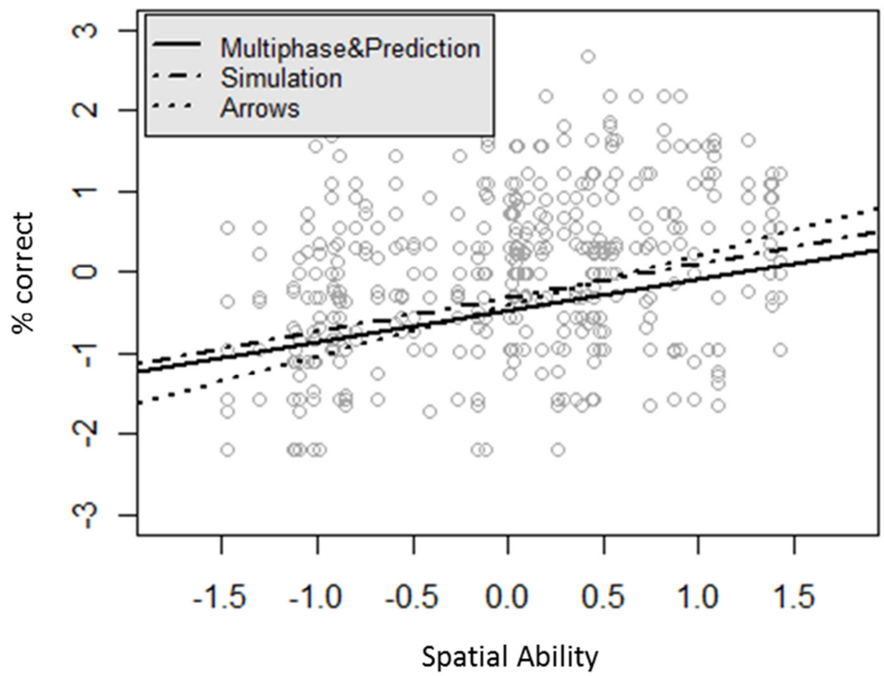


*Figure 14.* Non-significant interaction ($p$ = .10) between spatial ability and instructional conditions in Experiment two for post-test performance.

A second analysis examined the effect of VSWM (see Table 35). While controlling for verbal ability, neither VSWM nor its interactions with Format and Question Type accounted for a significant amount of variance in learning outcome.

Because the two measures of VSWM correlated only moderately ($r$ = .36), this null effect might have been due to poor validity of the composite VSWM measure. Composite VSWM was computed as the averaged of Symmetry Span and Rotation Span standardized scores. Either of the component measures alone might have been more valid than the composite measure. Therefore, hierarchical analyses were also performed on the individual measures. The pattern of results did not differ. That is, neither Symmetry Span nor Rotation Span accounted for significant variance in learning outcome after controlling for verbal ability. Neither measure interacted with instructional Format or Question Type.

Table 35. Results for comparisons between hierarchical models including VSWM and verbal ability in Experiment 2.

| Model | Terms in the Model* | log Likelihood | Likelihood Ratio | *P* |
|---|---|---|---|---|
| 1 | VSWM*Format*QuestionType | 130.77 | 3.11 | 0.21 |
| 2 | VSWM*Format | 129.21 | 0.54 | 0.76 |
| 3 | VSWM*QuestionType | 128.94 | 0.10 | 0.75 |
| 4 | VSWM | 128.89 | 0.35 | 0.55 |
| 5 | VerbalAbility*Format*QuestionType | 128.71 | | |
| | VerbalAbility*Format | | | |
| | VerbalAbility*QuestionType | | | |
| | VerbalAbility | | | |
| | Format*Topic*QuestionType | | | |
| | Topic*QuestionType | | | |
| | Format*QuestionType | | | |
| | Format*Topic | | | |
| | BackgroundKnowledge | | | |
| | QuestionType | | | |
| | Topic | | | |
| | Format | | | |

*Note:* Model #1 is the full model that includes all of the terms in the 2nd column. Each model includes the terms listed for that model, as well as all the terms below it. Likelihood ratios and p-values correspond to comparisons between the row on which they appear and the row beneath them.

### 3.3 Conclusion

Instructions to simulate motion, as well as the addition of arrows, were expected to improve learning from multiphase diagrams. However, the expected main effect of Format was not found ($p$ = .07), although the means in the three conditions trended in the expected direction. In contrast to this null result, the significant correlation between post-test scores and subjective ratings of motion visualizations indicated that the mental simulation activity was indeed effective in improving learning. When participants exerted more effort towards mental visualizations, their learning outcomes improved. It is possible that participants exerted high effort to mentally visualize motion regardless of the prompts; note that mean ratings in all three conditions exceeded 7.5 on a 10 point scale (see Table 23). This might have caused the lack of significant difference in ratings among conditions and also the lack of significant differences in learning outcomes among the conditions. For instructional designers, these results imply that visualization prompts might be superfluous if participants are sufficiently self-motivated to perform the visualizations. For learners, the results imply that visualizing motion might be beneficial when learning about physical systems.

The omnibus test of the interaction between spatial ability and instructional format did not show the expected effects. The arrows were expected to decrease the effect of spatial ability on learning, in comparison to the other conditions. Contrary to this expectation, the mean correlation between spatial ability and post-test performance in the arrows condition was higher than the mean correlations in the other two conditions (although the means were not significantly different). If these differences had been significant, the results would have been consistent with a pattern of enhancement, wherein spatial ability enhances learning from improved instructions. Participants with higher spatial ability might have been better able than those with lower spatial ability to use the arrows to construct mental models.

# CHAPTER 4: GENERAL DISCUSSION

Researchers and practitioners have long been interested in creating multimedia instructional materials that promote learning, retention, and transfer of knowledge. One topic of interest has been the design of instructional diagrams that convey information about dynamic physical systems (e.g., Hegarty, Kriz, & Cate, 2003; Mayer, 2001). Static instructional diagrams continue to be a prevalent format of instructional materials. Static diagrams do not directly convey information about dynamic motion, and therefore present a challenge to learners who attempt to construct dynamic mental models. Although animated instructions, which *do* present dynamic information, are becoming more common, instructional designers are likely to continue using static diagrams for some time. Therefore, research on ways to aid learning through static diagrams remains important.

Static diagrams cannot show motion explicitly, but depictive embellishments such as arrows might assist learners as they visualize motion. Learners with lower spatial ability might suffer a disadvantage when they attempt to perform mental animations, because their ability limits the accuracy of their animations. The present experiments examined instructional manipulations that were expected to interact with learners' spatial abilities. It was expected that the impact of spatial ability on learning outcome would be reduced by instructional manipulations that facilitated mental animations. Learning deficits for participants with lower spatial ability were expected to be reduced by environmental support for dynamic spatial reasoning.

Phase diagrams, depictive arrows, and problem solving (i.e., predictions) are three instructional factors that might facilitate spatial reasoning while learning. However, previous research has failed to disambiguate these manipulations (Hegarty & Just, 1993; Mayer & Gallini, 1990; Munzer, Seufert, & Brunken, 2009). The present experiments manipulated these instructional factors separately, with the expectation that they would impact learning about dynamic system behavior.

### 4.1 Acquisition of Behavioral and Structural Knowledge from Static Diagrams

A thorough understanding of how a dynamic physical system works is founded on not only knowledge of system structure, but also system movement. Because system movement is not directly depicted (i.e., animated) in static diagrams, learners might have difficulty constructing behavioral knowledge from static diagrams. Also, learners might need to construct knowledge about system components (i.e., their names and configuration) prior to constructing knowledge about the behavior of those structural components.

Contrary to expectations, participants performed better on behavior questions than structure questions in both experiments. It is possible that participants had retained accurate understandings of how the system behaved, but had forgotten many of the component names that were referenced in structure questions. If structural knowledge had been assessed by participants' ability to draw the systems from memory, rather than recall component names, then performance on structure questions might have exceeded performance on behavior questions in accordance with structure-behavior-function theory (Hmelo-Silver & Pfeffer, 2004). Structural knowledge might be easy to acquire when the structural information is overtly depicted, whereas behavioral knowledge might be more challenging to acquire if motion must be inferred.

### 4.2 Environmental Support for Learning

Learners could benefit from instructional manipulations that support essential cognitive processes. Multiphase diagrams are assumed to supply enough information for successful learning, but construction of a mental model is the learner's responsibility (Hegarty et al, 2003; Mayer & Gallini, 1990; Munzer et al, 2009). Learners must organize the depicted phases into a coherent mental model. Learners are sometimes able to construct mental models from single phase diagrams, but to do so they are required to perform more cognitive operations. Specifically, they must generate mental images of non-depicted states, and integrate those mental images with the depicted states to construct a complete mental model. Given the extra cognitive tasks required by single phase diagrams, which might be prone to error, it was expected that learning outcomes would differ significantly between single and multiphase diagram instructions.

Multiphase diagrams and arrows can be construed as forms of environmental support for learning. Morrow and Rogers (2008) identified two categories of environmental support: those that reduce task demands and those that support the use of cognitive resources. Both arrows and multiphase diagrams might reduce task demands. Arrows provide processing cues and externalize task elements, thereby alleviating the need to retain multiple motion vectors in working memory simultaneously. Multiphase diagrams, in comparison to single phase diagrams, also might reduce task demands, because learners do not need to visualize non-depicted states of the system when viewing multiphase diagrams. In contrast to multiphase diagrams and arrows, the prediction activity increases task demands and forces participants to rely on internal resources (i.e., cognitive abilities).

When the cognitive demands of a task are reduced, as might be the case with arrows and multiphase diagrams, one might expect cognitive resources such as spatial ability to be less influential on performance. This would represent a pattern of compensation, in which spatial ability compensates for the absence of environmental support. In contrast, environmental support that promotes the use of cognitive resources might improve learning only for those who have sufficient cognitive resources. This would represent a pattern of enhancement, in which spatial ability enhances the benefit of environmental support. The post-learning prediction activity might not qualify as a form of environmental support, but it does increase the demand on cognitive resources, and might therefore be expected to show a pattern of enhancement. The present experiments failed to support these hypotheses, possibly due to low statistical power. Future work on enhancement and compensation should systematically manipulate environmental supports while considering how the manipulations affect task demands and the use of cognitive resources.

The present experiments and previous studies have failed to show that multiphase diagrams enable significantly better learning than single phase diagrams. Similarly, studies have failed to show that depictive arrows improve learning. Previous studies comparing single phase diagrams to other formats have contained experimental confounds (Hegarty, et al., 2003; Mayer & Gallini, 1990; Munzer, et al., 2009). In each study, the single phase format differed by more than one aspect from the “improved” format. Both Mayer & Gallini (1990) and Munzer et al. (2009) compared single phase diagrams to multiphase diagrams with arrows. Learning improved in the latter condition, which might have been due to the multiphase diagrams, arrows, or both.

Results from Experiment 1 showed that multiphase diagrams alone did not improve learning significantly in the present sample, because participants did not perform significantly better in the multiphase condition than in the single diagram condition. When learning from multiphase diagrams, participants might have either failed to select information appropriately (e.g., look for differences between the phase diagrams), or they might have failed to organize the images into a coherent mental model.

The present experiments and previous studies have also failed to show that depictive arrows improve learning about physical systems. The effect size in Experiment 2 might have been too small to detect a significant improvement in the present sample. Although arrows might improve learning in future studies or in practice, it is not clear that the effect would be large. A larger, significant effect might be found if participants were asked to draw arrows rather than passively view them. Because successful learning is often an constructive process in which the learner generates knowledge as he or she processes information (Chi, 2009; Chi, Deleeuw, Chiu, & Lavancher, 1994; Mayer, 1997), it could be beneficial for participants to draw arrows on the diagrams as they are attempting to visualize motion. The overt, constructive action of drawing an arrow could stimulate and facilitate the covert, constructive action of simulating a mental model.

Despite the present null results, which might have been caused by insufficient statistical power, depictive arrows hold promise for improving learning. Depictive arrows are a form of environmental support that might help learners to understand the causal flow in a system. Other forms of environmental support, such as numbered labels, might have similar effects. For example, arrows or numbers could guide learners through the functional flow of software code as they are learning a programming language.

It was hoped that arrows would improve the visualization process for learners with lower spatial ability, but the data did not support this expectation. Other methods might be more effective in improving the visualization process for learners with lower spatial ability. Specifically, learners could be trained to visualize motion in diagrams. Prior studies on the transfer of visualization skills have yielded conflicting results, but a recent meta-analysis suggested that moderate transfer is attainable (Uttal, et al., in press). Thus, learners with lower spatial ability might benefit from practicing visualization skills prior to performing mental animations of diagrammatic instructions.

### 4.3 Benefits of Making Predictions

The prediction activity was expected to encourage participants to visualize motion (i.e., generate mental animations of the diagrams). Prior studies have not shown conclusively that predictions improve learning. Hegarty et al. (2003) compared single phase diagrams to multiphase diagrams with predictions, and found that learning improved in the latter condition. These results were ambiguous with regard to the separate effects of multiphase diagrams and predictions. Results from Experiment 1 showed that learning improved not from predictions alone, but from the combination of multiphase diagrams and predictions compared to single phase diagrams without predictions. Although this result was, again, ambiguous about whether predictions improved learning, there was a significant correlation between prediction performance and learning outcome. That is, participants who answered prediction questions correctly were more likely to answer post-test questions correctly. Therefore, it might be somewhat ineffective to simply ask participants to make predictions, but when participants make those predictions correctly, then their learning might improve.

In the present experiments and in Hegarty et al. (2003), the term "predictions" has referred to the activity of predicting how the system would behave under circumstances that were not described in the lessons. The prediction activity occurred after participants were exposed to the learning materials, rather than before. In other studies (e.g., Kasmer & Kim, 2012), the term "prediction" has referred to activities performed before lesson presentation. In either case, learners have benefited from making predictions, regardless of the time at which they were made. However, the reasons for the benefits might differ. Kasmer and Kim (2012) argued that pre-lesson predictions might have provoked interest and alerted learners about their knowledge deficiencies that they should seek to remediate while learning. Thus, pre-lesson predictions were assumed to affect students' processing of the lesson material. In contrast, post-lesson predictions are assumed to exert their effect by invoking mental model simulation (i.e., "mental animation"), which can be seen as a form of practice. When mental model simulation is practiced immediately after learning, the model is likely to be more robust to decay over time. The choice of whether predictions should be made before or after lesson presentation might depend on the complexity of the system and the prior knowledge of the learner. When learners have low prior knowledge of a complex system, pre-lesson predictions might be fruitless.

The act of performing mental animations while answering prediction questions, if done correctly, was expected to improve the formation of a mental model. Consistent with this hypothesis, learning outcomes were correlated with performance on the prediction questions. High spatial ability, which was also correlated with prediction performance, might have helped participants to correctly simulate motion while answering the prediction questions. If these accurate mental simulations led to better learning, then one would expect prediction performance to mediate the relationship between spatial ability and learning outcome. Indeed, the effect of spatial ability on learning outcome was partially mediated by prediction performance. The mediation effect suggested that spatial ability improved mental animations during the prediction activity, which in turn led to better learning.

The results of the mediation analysis showed that spatial ability had a direct effect on learning in the multiphase-plus-prediction condition, as well as an additional mediated effect that was dependent upon performance on the prediction questions. A significant portion of spatial ability's effect on learning was mediated by spatial ability's effect on prediction performance. Thus, spatial ability enhanced the benefit that participants derived from the prediction activity. The enhancement might have been due to better mental model formation during the predictions activity, followed by better performance on the post-test.

This interpretation of the enhancing effect of spatial ability would have been further supported by a significant interaction between spatial ability and instructional format. Specifically, if spatial ability had been more predictive of learning from the condition with predictions than those without predictions, then the difference could have been due to spatial ability's enhancement of the prediction activity. However, this expected interaction was not significant ($p$ = .054). Nonetheless, the significant mediation effect indicated that poor prediction performance was part of the reason why learners with lower spatial ability did not learn as well as those with higher spatial ability in the multiphase-plus-prediction condition.

### 4.4 Active Learning from Optimal Instructions

The efficacies of instructional manipulations are partially dependent upon how learners process the presented information. It is the learner's responsibility to direct attention and cognitive resources appropriately, although instructional designs can facilitate this. In the Cognitive Theory of Multimedia Learning, Mayer (1997) proposed three distinct stages of knowledge construction from multimedia instructions, in which information is first selected by the learner from the presented materials and then organized into a coherent mental representation (see Figure 2). These two stages occur in the verbal processing pathway and the pictorial processing pathway. In the final step, the verbal and pictorial mental models are integrated. The manipulations in the present experiments might have directly affected the first two processing stages (i.e., selection and organization). Arrows and phase diagrams are examples of factors that are external to the learner, and they might affect how learners select information. Specifically, arrows might be interpreted as cues about which system components should be selected from a given diagram for further cognitive processing. Similarly, differences between the system phases, as depicted in successive diagrams, might cue learners to select components that have changed between the two diagrams (i.e., if a component's position changed from one diagram to the next, then the learner might have selected that component for further processing). If differences between successive diagrams were not salient, then learners might have failed to select the appropriate components. When information is not selected, then it cannot affect the subsequent stages of mental model development.

In the present studies, the inefficacy of arrows and multiphase diagrams might have been caused by learners' inability to appropriately select diagrammatic elements for further cognitive processing. Learners might need cues to guide their selection of diagrammatic elements. However, appropriate *selection* of information does not guarantee appropriate *processing* of that information in the subsequent organization and integration stages. Learning from multimedia instructions is a generative process in which learners actively construct knowledge, not only by selecting information from the materials, but also by actively organizing and integrating that information to construct a mental model (Mayer, 1997; Mayer & Moreno, 2003).

Each stage of learning from multimedia instructions can be affected by environmental support and the learner's cognitive abilities. Multimedia instructions should be designed to promote the *selection* of information at appropriate times, the *organization* of information into a dynamic mental representation, and the *integration* of that representation with existing knowledge and other presented information. If an instructional design neglects any one of these three stages, then the effectiveness of the instructions could be limited. Moreover, an instructional design could fail to promote learning if it demands cognitive resources that exceed the learner's abilities and capacities. Learners with lower spatial abilities might need more environmental support for cognitive processes such as mental animation than those with higher spatial abilities. Learning might be improved for all learners, regardless of spatial ability, by instructions that both encourage learners to simulate nascent mental models and provide appropriate environmental support for those simulations.

## APPENDIX A:

## BACKGROUND QUESTIONNAIRE

What is your gender? Male Female

Please report your SAT scores:

Verbal (critical reading) _____ Mathematics _____ Writing _______

Please rate your knowledge on the following items (circle a number):

1. General knowledge about how valves work in various machines

*(Low)* 1 2 3 4 5 6 7 8 9 10 *(high)*

2. Knowledge of how dual-action pumps work

*(Low)* 1 2 3 4 5 6 7 8 9 10 *(high)*

3. General knowledge about how automobiles work

*(Low)* 1 2 3 4 5 6 7 8 9 10 *(high)*

4. Knowledge of how carburetors work

*(Low)* 1 2 3 4 5 6 7 8 9 10 *(high)*

5. General knowledge of plumbing

*(Low)* 1 2 3 4 5 6 7 8 9 10 *(high)*

6. Knowledge of how toilet tanks work

*(Low)* 1 2 3 4 5 6 7 8 9 10 *(high)*

## APPENDIX B:

## LESSON MATERIALS FOR THE PUMP TOPIC - SINGLE DIAGRAM CONDITION

The single diagram condition consisted of five pages as follows:

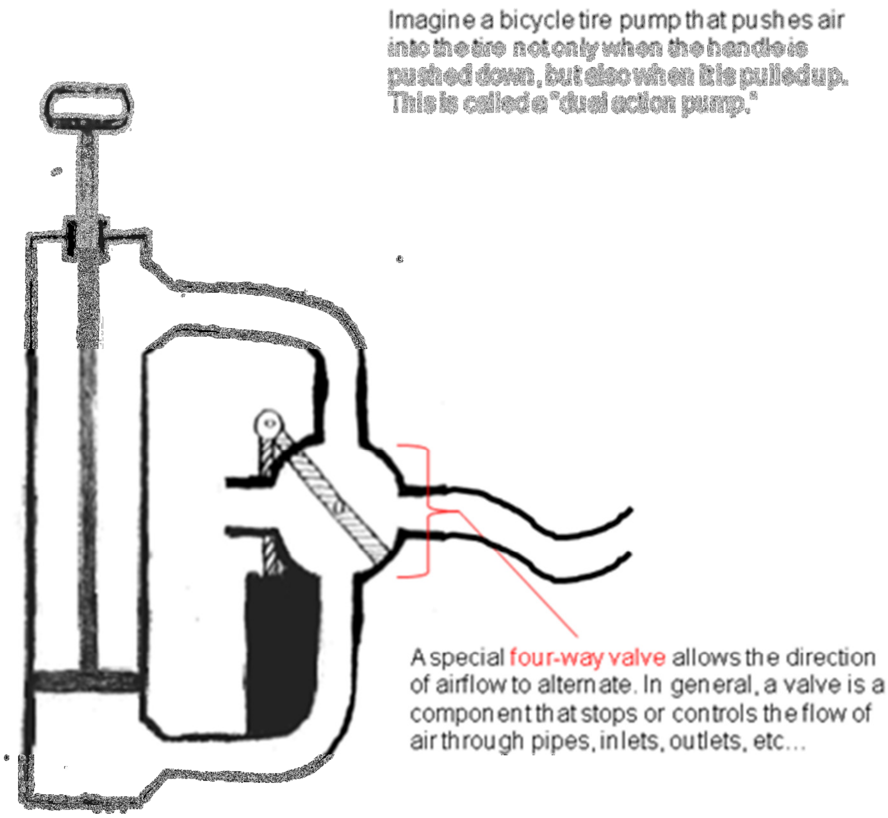


Figure 15. Single diagram lesson for the dual-action air pump, page 1.

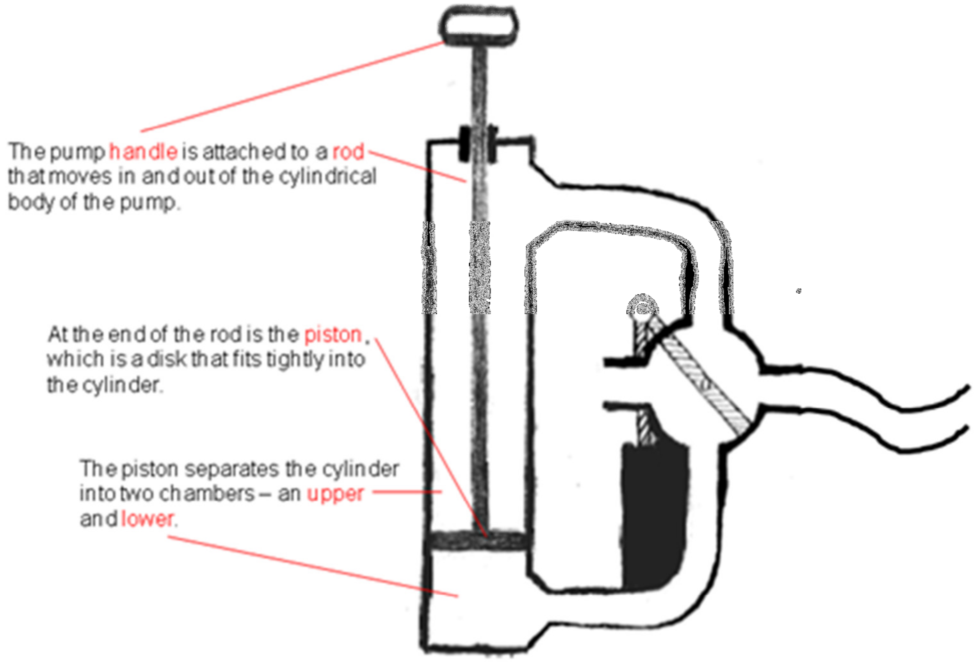


Figure 16. Single diagram lesson for the dual-action air pump, page 2.

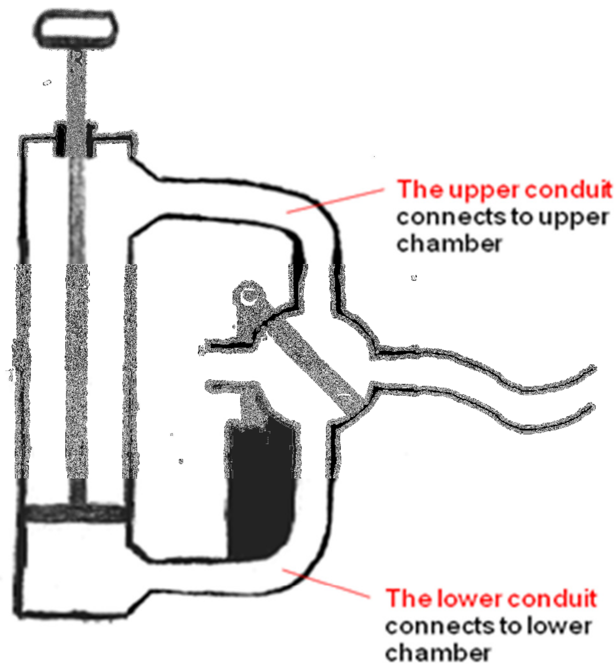


Figure 17. Single diagram lesson for the dual-action air pump, page 3.

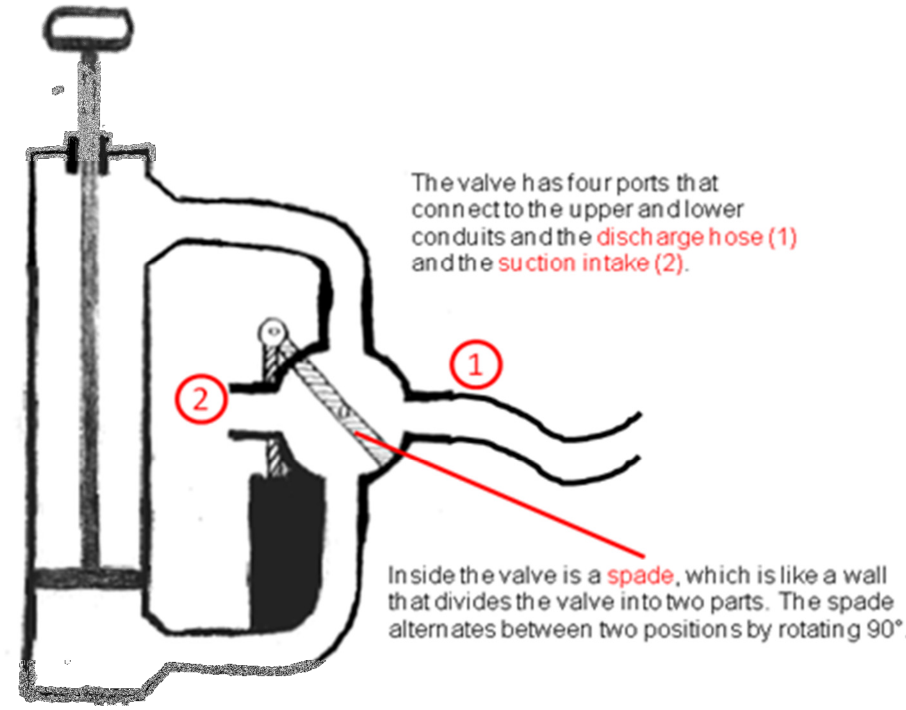


Figure 18. Single diagram lesson for the dual-action air pump, page 4.

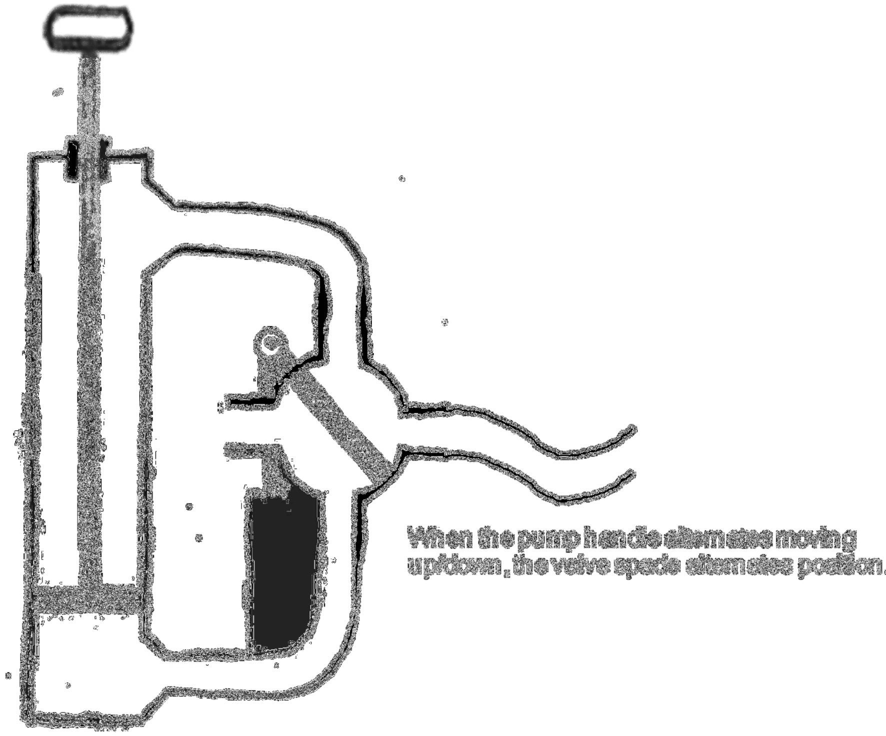


Figure 19. Single diagram lesson for the dual-action air pump, page 5.

## APPENDIX C:

## LESSON MATERIALS FOR THE PUMP TOPIC – MULTIPHASE DIAGRAMS

Instructions for the multiphase diagrams condition and the multiphase-plus-predictions condition consisted of seven pages, as follows:

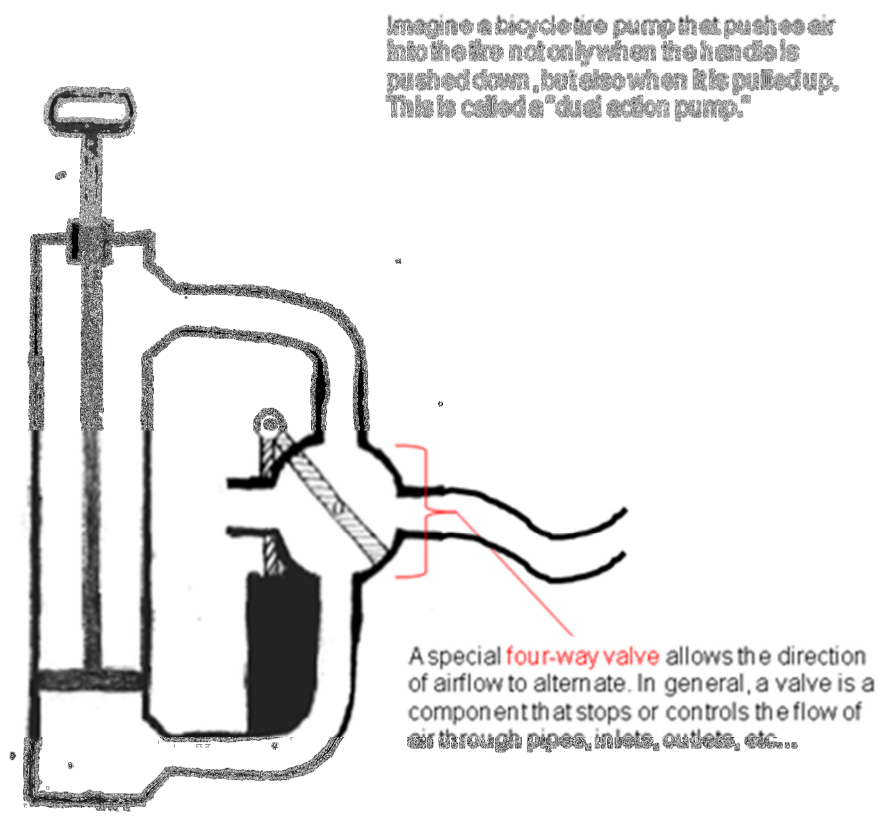


Figure 20. Multiphase diagram lesson for the dual-action air pump, page 1.

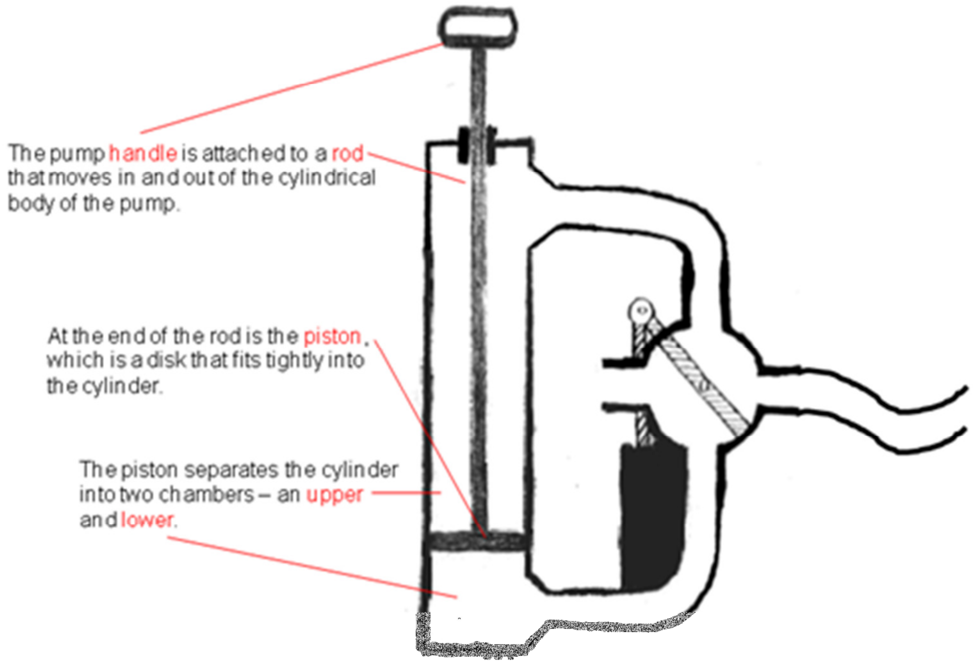


Figure 21. Multiphase diagram lesson for the dual-action air pump, page 2.

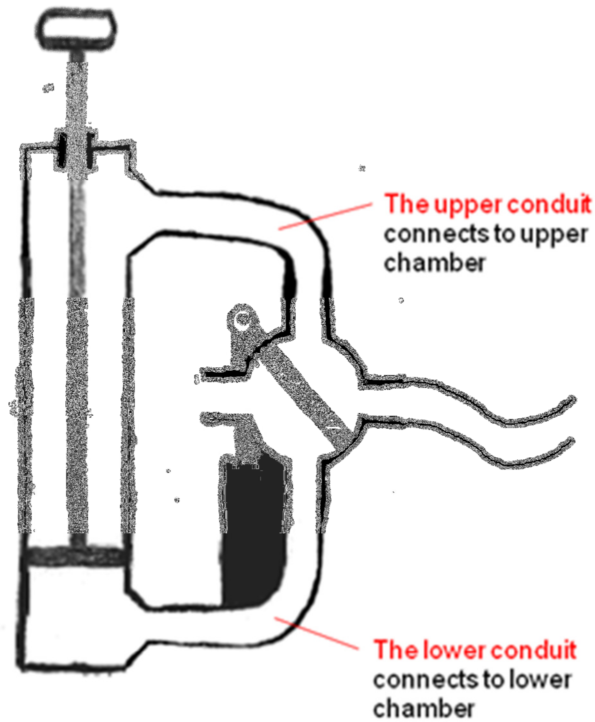


Figure 22. Multiphase diagram lesson for the dual-action air pump, page 3.

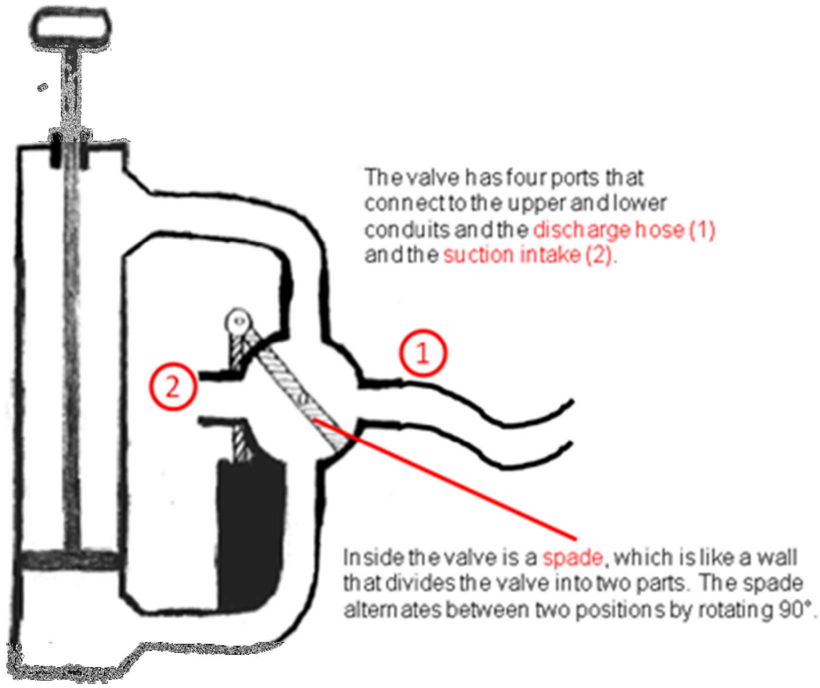


Figure 23. Multiphase diagram lesson for the dual-action air pump, page 4.

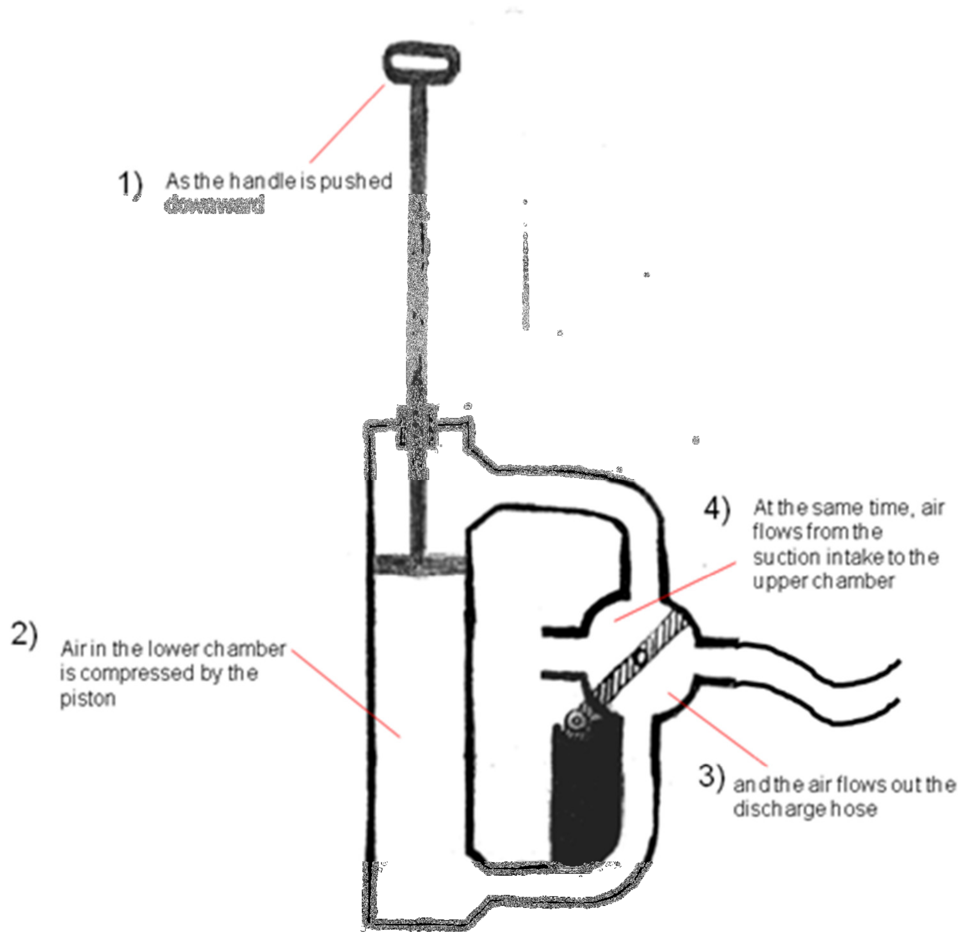


Figure 24. Multiphase diagram lesson for the dual-action air pump, page 5.

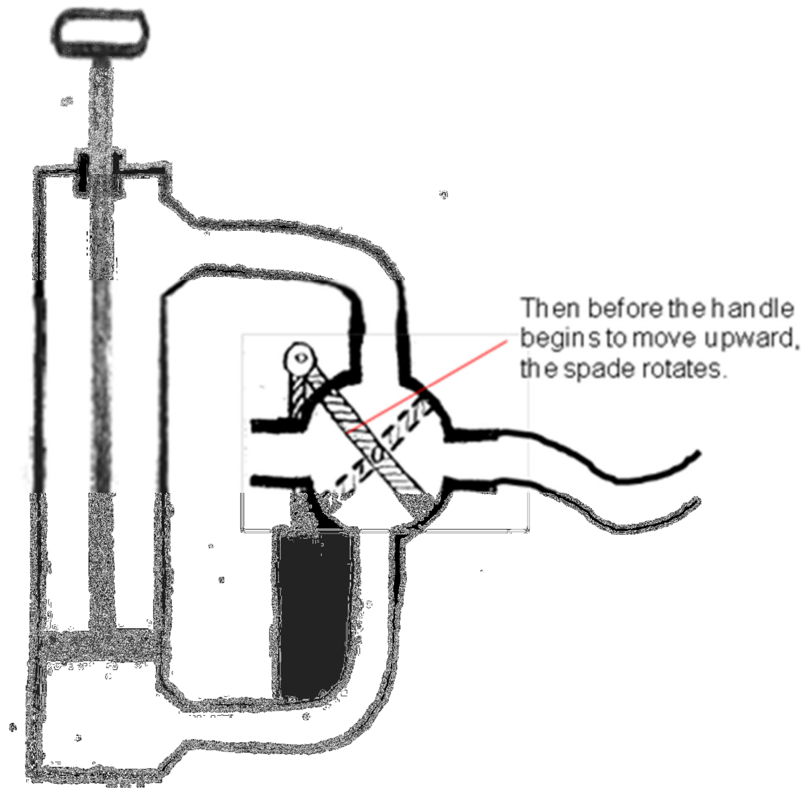


Figure 25. Multiphase diagram lesson for the dual-action air pump, page 6.

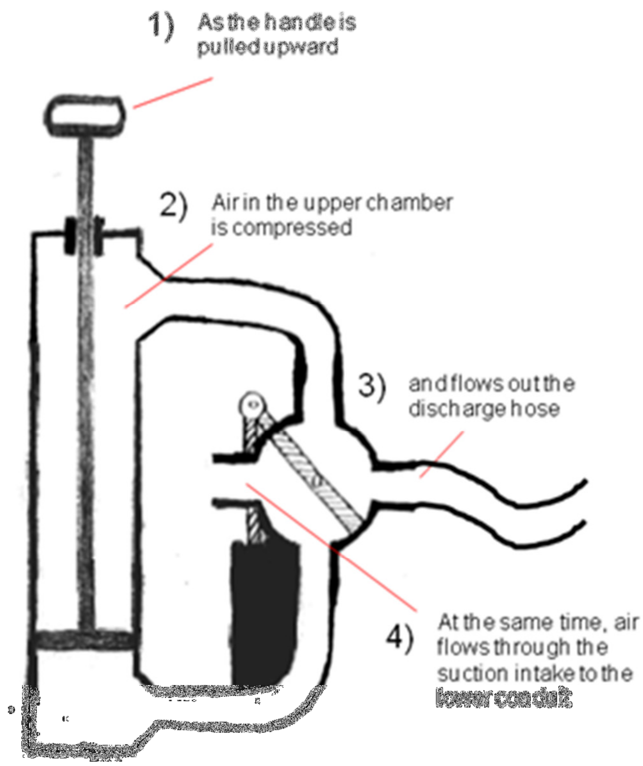


Figure 26. Multiphase diagram lesson for the dual-action air pump, page 7.

## APPENDIX D:

## LEARNING ASSESSMENT QUESTIONS FOR THE PUMP TOPIC

This document contains examples of assessment questions for the pump topic only. Comparable questions were used for the other two topics (i.e., carburetors and toilet tanks).

Answers are shown in bold italics. For some items multiple answers were acceptable; these were selected by pilot testing. The first eight questions assess structural knowledge.

1. The ____***rod/shaft/bar***__________ connects the handle to the piston.

2. The pump chamber is divided by the ______***piston***_______________.

3. How many ports does the valve have?

   *Four*

4. The suction intake port is directly opposite the __***discharge/outtake/output/outlet***____ port.

5. How many different stopped positions does the valve spade stop in?

   *Two*

6. The _______***suction intake***_________ port allows air to enter the system.

7. How many chambers does the pump have?

   *Two*

8. What part of the pump is located at the lower end of the upper conduit?

   *Valve*

The following set of 14 questions assesses knowledge of pump system behavior.

1. As air leaves the pump, it passes *last* through the ***__discharge hose* or *outtake/output/outlet_***.

2. When air exits the upper chamber as the piston moves upward, the air passes first into the ***______upper conduit__________***.

4. When the valve spade rotates, how far does it rotate (i.e., how many degrees)?

*Ninety*

5. How could the pump be modified to make it fill a bicycle tire more rapidly without moving the handle faster?

*Make the chamber wider (larger diameter)*

6. If an operator reaches into the valve to clean it while the pump is being operated, his or her hand could be injured. How?

*The valve spade moves back & forth past the openings, so it could trap or sever fingers*

7. How could the valve spade's rotation be modified to prevent injury?

*Make the ends of the spade sweep across the conduit ports instead of the intake/discharge ports*

8. When air exits the lower chamber as the piston moves downward, the exiting air passes first into the ***____lower conduit_____________***.

9. If the pump were damaged, causing a large hole to open in the wall of the upper conduit, could the pump still push air out through the discharge hose? *Why or why not*?

*Yes, the lower chamber would work. The pump would function like a single-action pump instead of dual-action.*

10. As the piston moves upward, the *volume* of air in the *upper conduit*

  a) increases.

  b) decreases.

  ***c) does not change.***

11. As the piston moves upward, the air *pressure* in the *lower chamber*

a) increases.

***b) decreases.***

c) does not change.

12. How could the pump be modified so that two people could simultaneously operate it, effectively doubling the force on the piston. Imagine that the rod cannot bear more force than that generated by one operator, and the rod cannot be modified.

  ***add a rod to the other side of the piston*** or ***use two rods***

13. Imagine that a balloon is attached to the discharge hose. If the valve spade failed to rotate 90° when the motion of the piston reversed, what would happen to the balloon as the pump handle moved up and down?

  *It would inflate and deflate repeatedly*

14. If air flow in the upper conduit were completely blocked by an obstruction, could the pump still be operated? *Why or why not*?

***No, because the trapped air in the upper chamber/conduit would prevent the piston from moving upwards because the air pressure would be too great.***

***(If "No" without an explanation, then give 0.5 points).***

***<u>OR</u>: Yes, it could be operated with excessive force, because the air in the upper chamber could be compressed/expanded.***

**APPENDIX E:**

**LESSON MATERIALS FOR THE PUMP TOPIC – MULTIPHASE DIAGRAMS WITH ARROWS**

This condition (arrows) appeared only in Experiment 2. The pump lesson consisted of seven slides.

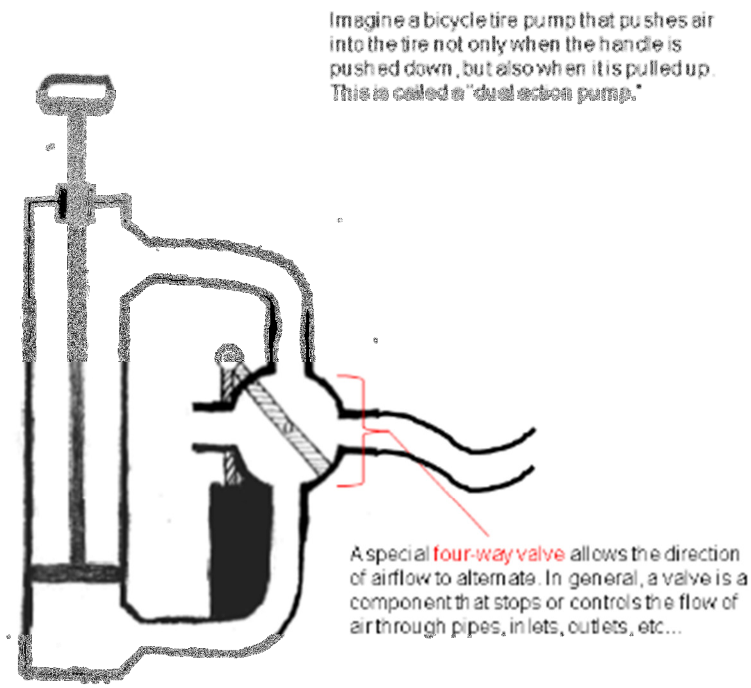


Figure 27. Multiphase diagram lesson with arrows for the dual-action air pump, page 1.

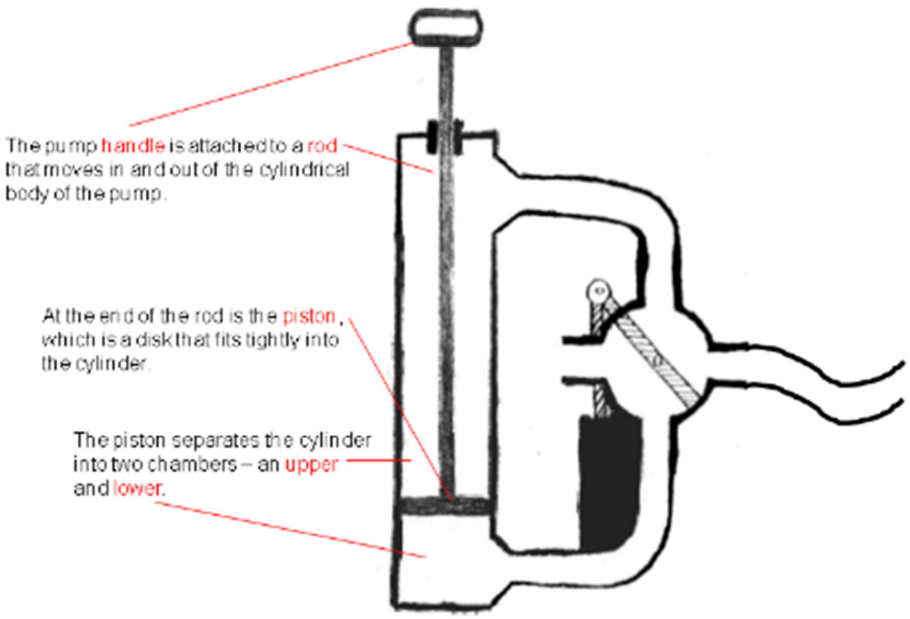


Figure 28. Multiphase diagram lesson with arrows for the dual-action air pump, page 2.

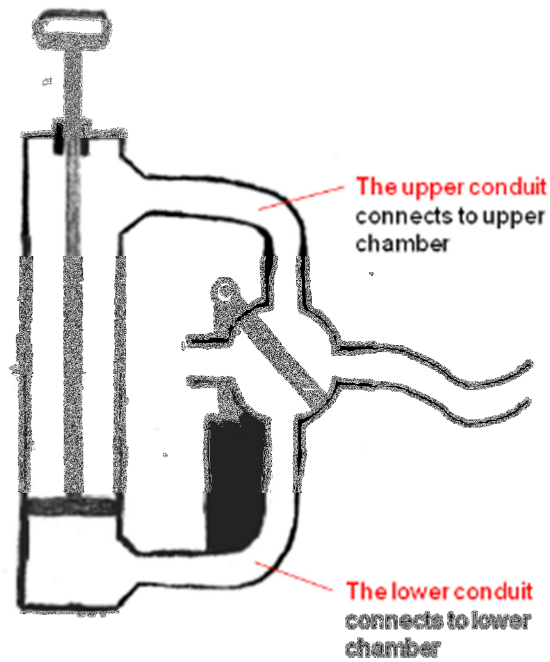


Figure 29. Multiphase diagram lesson with arrows for the dual-action air pump, page 3.

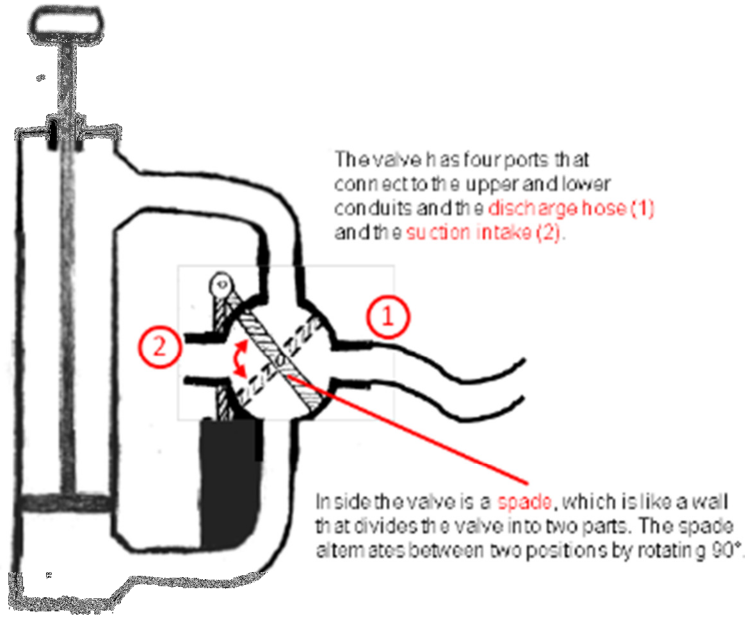


Figure 30. Multiphase diagram lesson with arrows for the dual-action air pump, page 4.

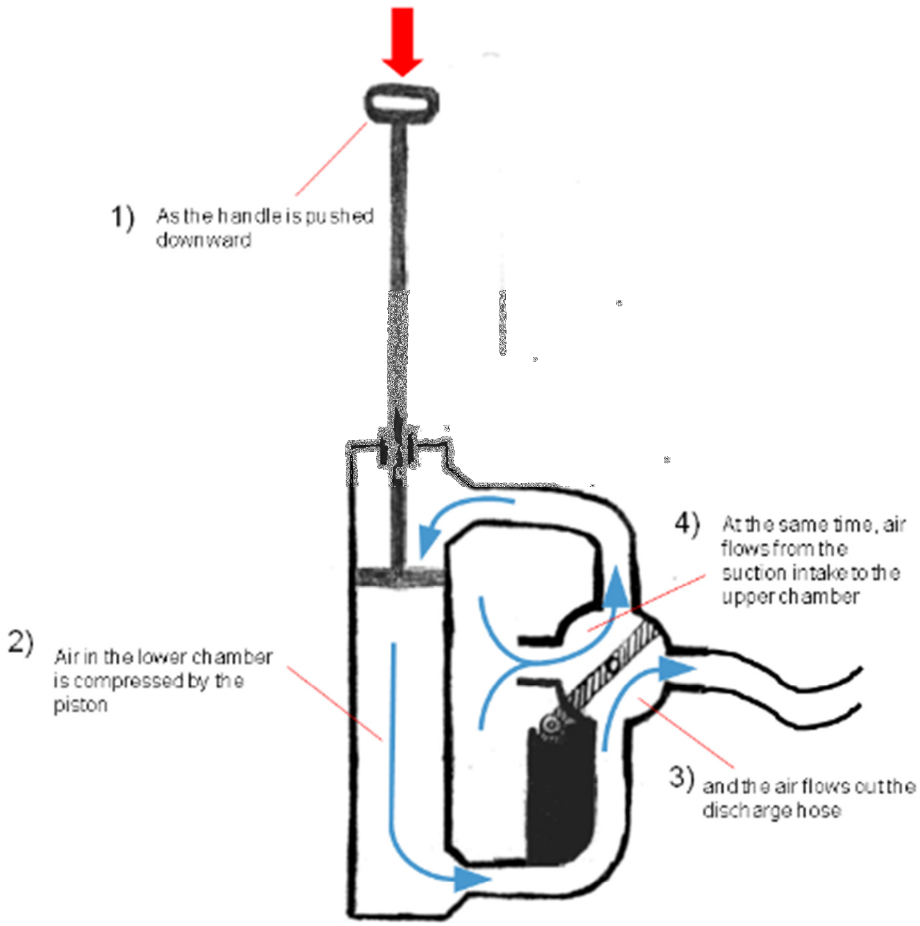


Figure 31.  Multiphase diagram lesson with arrows for the dual-action air pump, page 5.

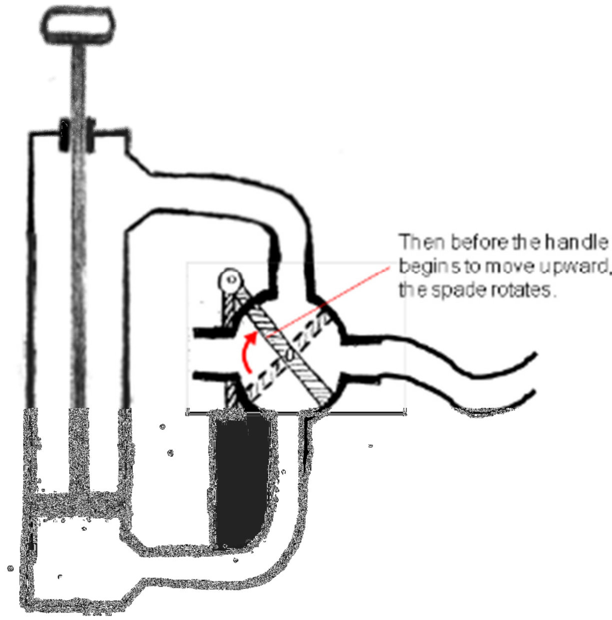


Figure 32. Multiphase diagram lesson with arrows for the dual-action air pump, page 6.

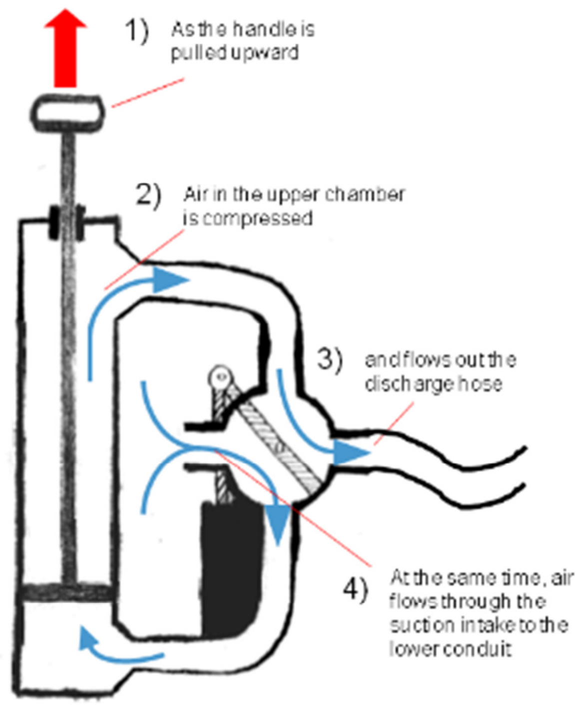


Figure 33. Multiphase diagram lesson with arrows for the dual-action air pump, page 7.

## APPENDIX F:

## TEXT PASSAGE FOR THE CLOZE TEST

This passage was derived from the opening pages of Oliver Twist (Dickens, 1867/2003). Small modifications were made to replace archaic language. For example, the phrase "take upon himself the *office* of respiration" was changed to "take upon himself the *task* of respiration."

Although I am not disposed to say that being born in a workhouse is the most fortunate thing that can possibly happen to a human being, I do mean to say that _____in______this particular instance, it was the __ __best__ _ thing for Oliver Twist that could __possibly__ _have occurred. The fact is, that ____there___was considerable difficulty in inducing Oliver _____to______take upon himself the task of __respiration____--an activity necessary for our existence; _____and_____for some time he lay gasping _____on______a little mattress, rather unequally balanced __between___this world and the next: the __balance___being decidedly in favor of the ___latter____. During this brief period, if Oliver ____had______been surrounded by careful grandmothers, anxious ___aunts_____, experienced nurses, and doctors with great ____skill____, he would most inevitably and indubitably ___have_____been killed in no time. There ___was_____ nobody by, however, but a poor ____old______woman, who was rendered rather drunk ____by______a large quantity of beer, and _____a_____parish surgeon who did such matters ____for_____a fee. Oliver and Nature fought ____out_____the point between the two onlookers. ___The______result was, that, after a few __struggles_ _, Oliver breathed, sneezed, and proceeded to ___ _cry_____to the other occupants of the _ workhouse_ _the fact that a new burden ____was_____now imposed upon the parish, by _ _letting___out as loud a cry as ___could___have been expected from a male ____infant___who had not long possessed that ____very_____useful appendage–a voice—for ___much_____ longer space of time than three ____minutes____and a fifteen seconds.

As Oliver ___gave____this first proof of the proper ___action____of his lungs, the patchwork blanket ___that_____was carelessly flung over the iron ____bed______ rustled. The pale face of a ____young_____woman was raised feebly from the ___pillow_____, and a faint voice imperfectly articulated ____the_____words, 'Let me see my child, ____and_____die.'

The surgeon had been sitting ____with______his face turned towards the fire, __warming___and rubbing the palms of his ____hands_____. As the young woman spoke, he ____rose______ and advanced to the bed's head, ____saying_____ with more kindness than might have _been____expected of him:

'Oh, you must ____not_____talk about dying yet.'

## REFERENCES


Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2000). Explorations of crystallized intelligence - Completion tests, cloze tests, and knowledge. *Learning and Individual Differences, 12*(1), 105-121.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*(1), 30-60.

Ayres, P., & Paas, F. (2007). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21*(6), 695-700.

Bacon, A. M., Handley, S. J., & Newstead, S. E. (2003). Individual differences in strategies for syllogistic reasoning. *Thinking & Reasoning, 9*(2), 133-168.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist, 7*(2), 85-97.

Baron, R. M., & Kenny, D. A. (1986). The moderator mediator variable distinction in social psychological research - conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Bartholome, T., & Bromme, R. (2009). Coherence Formation When Learning From Text and Pictures: What Kind of Support for Whom? *Journal of Educational Psychology, 101*(2), 282-293.

Betrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning*. Cambridge: Cambridge University Press.

Betrancourt, M., Dillenbourg, P., & Clavien, L. (2008). Display of key pictures from animation: Effects on learning. In J. Rouet, R. Lowe & W. Schnotz (Eds.), *Understanding multimedia documents*. New York: Springer Science + Business Media.

Bhatta, S. R., & Goel, A. (1997). Learning generic mechanisms for innovative strategies in adaptive design. [Article]. *Journal of the Learning Sciences, 6*(4), 367-396.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98*(1), 182-197.

Byrne, M. D., Catrambone, R., & Stasko, J. T. (1999). Evaluating animations as student aids in learning computer algorithms. *Computers & Education, 33*(4), 253-278.

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension - psycholinguistic processing model of verification [Article]. *Psychological Review, 82*(1), 45-73.

Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor analytic studies* Cambridge Cambridge University Press

Catrambone, R. (1995). Aiding subgoal learning: Effects on transfer. *Journal of Educational Psychology 87*(1), 5-17.

Catrambone, R., & Seay, A. F. (2002). Using animation to help students learn computer algorithms. [Article]. *Human Factors, 44*(3), 495-511.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instructions. *Cognition and Instruction, 8*(4), 293-332.

Chi, M. T. H. (2009). Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Topics in Cognitive Science, 1*, 73-105.

Chi, M. T. H., Deleeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves learning. [Article]. *Cognitive Science, 18*(3), 439-477.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25*(2), 315-324.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Cortina, J. M. (1993). What is coefficient alpha - an examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals, and how to read pictures of data. *American Psychologist, 60*, 170-180.

de Koning, B. B., Tabbers, H. K., Rikers, R., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology, 21*(6), 731-746.

Dickens, C. (1867/2003). *Oliver Twist*. New York: Fine Creative Media, Inc.

Dutke, S., & Rinck, M. (2006). Multimedia learning: Working memory and the learning of word and picture diagrams. *Learning and Instruction, 16*(6), 526-537.

Egan, D. E., & Grimes-Farrow, D. D. (1982). Differences in mental representations spontaneously adopted for reasoning *Memory & Cognition, 10*(4), 297-307.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

Eliot, J., & Smith, I. M. (1983). *An International Directory of Spatial Tests*. Windsor, Great Britain: Nfer-Nelson.

Faraday, P., & Sutcliffe, A. (1997). *Designing effective multimedia presentations.* Paper presented at the Proceedings of the ACM CHI '97.

Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction, 20*(3), 216-224.

Ford, M. (1995). Two modes of mental representation and problem solution in syllogistic reasoning [Article]. *Cognition, 54*(1), 1-71.

Gentner, D., & Stevens, A. L. (1983). *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*(6), 511-525.

Gyselinck, V., Cornoldi, C., Dubois, V., De Beni, R., & Ehrlich, M. F. (2002). Visuospatial memory and phonological loop in learning from multimedia. *Applied Cognitive Psychology, 16*(6), 665-685.

Gyselinck, V., Ehrlich, M. F., Cornoldi, C., de Beni, R., & Dubois, V. (2000). Visuospatial working memory in learning from multimedia systems. *Journal of Computer Assisted Learning, 16*(2), 166-176.

Gyselinck, V., & Tardieu, H. (1999). The role of illustrations in text comprehension: what, when, for whom, and why? In H. van Oostendorp & S. R. Goldman (Eds.), *The Construction of Mental Representations During Reading*. Mahwah, NJ: Lawrence Erlbaum Associates.

Halasz, F. G. (1984). *Mental models and problem solving in using a calculator.* Stanford University.

Hegarty, M. (1992). Mental animation - inferring motion from static displays of mechanical systems *Journal of Experimental Psychology-Learning Memory and Cognition, 18*(5), 1084-1102.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*(6), 717-742.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction, 21*(4), 325-360.

Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition, 22*(4), 411-430.

Hegarty, M., & Steinhoff, K. (1997). Individual differences in use of diagrams as external memory in mechanical reasoning. *Learning and Individual Differences, 9*(1), 19-42.

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science, 30*(3), 581-592.

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. [Article]. *Cognitive Science, 28*(1), 127-138.

Hoffler, T. N., & Leutner, D. (2010). The role of spatial ability in learning from instructional animations - Evidnce for an ability-as-compensator hypothesis. *Computers in Human Behavior.*

Huk, T. (2006). Who benefits from learning with 3D models? the case of spatial ability. *Journal of Computer Assisted Learning, 22*(6), 392-404.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science, 25*(1), 51-71.

Jamet, E., Gavota, M., & Quaireau, C. (2008). Attention guiding in multimedia learning. *Learning and Instruction, 18*(2), 135-145.

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*(2), 137-172.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(4), 351-371.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology-General, 133*(2), 189-217.

Kasmer, L. A., & Kim, O. K. (2012). The nature of student predictions and learning opportunities in middle school algebra. *Educational Studies in Mathematics, 79*(2), 175-191.

Kieras, D. E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science, 8*, 255-273.

Kim, H., & Forbus, K. D. (1993). *Qualitatve reasoning about fluids and mechanics*: University Press.

Kline, K. A., & Catrambone, R. (2009). The influence of spatial ability on multimedia learning. *Proceedings of the Human Factors and Ergonomics Society, 53*, 1249-1253.

Kline, K. A., & Catrambone, R. (2011). Learning from Multiphase Diagrams: Effects of Spatial Ability and Visuospatial Working Memory Capacity. *Proceedings of the Human Factors and Ergonomics Society, 55*.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information - evidence from studies of image scanning [Article]. *Journal of Experimental Psychology-Human Perception and Performance, 4*(1), 47-60.

Lee, H. (2007). Instructional design of web-based Simulations for learners with different levels of spatial ability. *Instructional Science, 35*(6), 467-479.

Liu, L., & Hmelo-Silver, C. E. (2009). Promoting Complex Systems Learning through the Use of Conceptual Representations in Hypermedia. [Article]. *Journal of Research in Science Teaching, 46*(9), 1023-1040.

Lohman, D. F. (1996). Spatial ability and *g*. In I. Dennis & P. Tapsfield (Eds.), *Human Abilities: their Nature and Measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Macleod, C. M., Hunt, E. B., & Mathews, N. N. (1978). Individual differences in the verification of sentence-picture relationships *Journal of Verbal Learning and Verbal Behavior, 17*(5), 493-507.

Mayer, R. E. (1981). The psychology of how novices learn computer programming. *Computing Surveys, 13*(1), 121-141.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge: Cambridge University Press.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth 10000 words. *Journal of Educational Psychology, 82*(4), 715-726.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology-Applied, 11*(4), 256-265.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth 1000 words - extensions of dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389-401.

Meneghetti, C., Gyselinck, V., Pazzaglia, F., & De Beni, R. (2009). Individual differences in spatial text processing: High spatial ability can compensate for spatial working memory interference. [Article]. *Learning and Individual Differences, 19*(4), 577-589.

Mikk, J. (2008). Sentence length for revealing the cognitive load reversal effect in text comprehension. *Educational Studies, 34*(2), 119-127.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, P. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology-General, 130*(4), 621-640.

Morrow, D. G., & Rogers, W. A. (2008). Environmental support: An integrative framework. *Human Factors, 50*(4), 589-613.

Munzer, S., Seufert, T., & Brunken, R. (2009). Learning from multimedia presentations: Facilitation function of animations and spatial abilities. *Learning and Individual Differences, 19*(4), 481-485.

Narayanan, N. H., & Hegarty, M. (2002). Multimedia design for communication of dynamic information. *International Journal of Human-Computer Studies, 57*(4), 279-315.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.).* New York: McGraw-Hill.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford Univesity Press.

Payne, S. J. (1991). A descriptive study of mental models. *Behaviour and Information Technology, 10*(1), 3-21.

Pellegrino, J. W., Alderton, D. L., & Shute, V. J. (1984). Understanding spatial ability. *Educational Psychologist, 19*(3), 239-253.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test - different versions and factors that affect performance *Brain and Cognition, 28*(1), 39-58.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Pinker, S., & Kosslyn, S. M. (1978). The representation and manipulation of three-dimensional space in mental images. *Journal of Mental Imagery, 2*, 69-84.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*(2), 221-243.

Schnotz, W. (2005). An Integrated Model of Text and Picture Comprehension. In R. E. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning*. Cambridge: Cambridge University Press.

Schnotz, W., & Lowe, R. (2003). External and internal representations in multimedia learning - Introduction. *Learning and Instruction, 13*(2), 117-123.

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science, 20*(4), 457-497.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology-General, 125*(1), 4-27.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of 3-dimensional objects *Science, 171*(3972), 701-&.

Smith, E. E., & Goodman, L. (1984). Understanding written instructions: The role of an explanatory schema. *Cognition and Instruction, 1*(4), 359-396.

Staggers, N., & Norcio, A. F. (1993). Mental models: concepts for human-computer interaction research. *International Journal of Man-Machine Studies, 38*, 587-605.

Stumpf, H., & Eliot, J. (1999). A structural analysis of visual spatial ability in academically talented students. [Article]. *Learning and Individual Differences, 11*(2), 137-151.

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review, 22*(2), 123-138.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185-233.

Taylor, W. L. (1957). Cloze readability scores as indexes of individual differences in comprehension and aptitude *Journal of Applied Psychology, 41*(1), 19-26.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies, 57*(4), 247-262.

Uttal, D. H., Meadow, N. G., Hand, L. L., Lewis, A. R., Warren, C., & Newcombe, N. (in press). Training Spatial Skills: What Works, for Whom, and for How Long? *Psychological Bulletin*.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations - a group test of 3-dimensional spatial visualization *Perceptual and Motor Skills, 47*(2), 599-604.

Williams, M. D., Hollan, J. D., & Stevens, A. L. (1983). Human reasoning about a simple physical system. In A. L. Stevens & D. Gentner (Eds.), *Mental Models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wilson, J. R., & Rutherford, A. (1989). Mental models: Theory and application in human-factors. *Human Factors, 31*(6), 617-634.

Winer, B. J. (1971). *Statistical Principles in Experimental Design*. New York: McGraw-Hill.

Yang, E. M., Andre, T., & Greenbowe, T. J. (2003). Spatial ability and the impact of visualization/animation on learning electrochemistry. *International Journal of Science Education, 25*(3), 329-349.